



Swansea University  
Prifysgol Abertawe



## Swansea University E-Theses

---

# Narrow content in the philosophy of mind.

Davies, Mark

### How to cite:

---

Davies, Mark (2009) *Narrow content in the philosophy of mind..* thesis, Swansea University.  
<http://cronfa.swan.ac.uk/Record/cronfa43049>

### Use policy:

---

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence: copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder. Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

Please link to the metadata record in the Swansea University repository, Cronfa (link given in the citation reference above.)

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

# **NARROW CONTENT IN THE PHILOSOPHY OF MIND**

**STUDENT – MARK DAVIES**

**SUPERVISOR – PROFESSOR STEVE EDWARDS**

**SUBMITTED TO THE UNIVERSITY OF WALES IN  
FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY OF HEALTH SCIENCE**

**SWANSEA UNIVERSITY**

**2009**

ProQuest Number: 10821439

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10821439

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346





## SUMMARY

My thesis examines the notion of narrow content in the philosophy of mind. Narrow content is defined as a type of mental state that is shared between internal physical duplicates in Twin Earth-type thought experiments. In these thought experiments, changes in the physical and/or social environments of the duplicates are usually taken to have the result that the contents of certain of the duplicates' mental states, e.g., beliefs, etc., are different. The upshot of this is that internalism with regards to mental content seems to be refuted, as the duplicates concerned share an identity of internal physical properties, then if internalism about mental content is true, their mental states should also be content-identical. However, despite this possibility, it seems to me that there is still a strong intuition that the duplicates in these Twin Earth-type situations do share type-identical mental states which can be individuated by the states' narrow content. The aim of the thesis is to examine several of the most popular construals of narrow content in the literature, to ascertain whether there is a construal that provides an adequate narrow content. To help with this task I suggest three conditions of adequacy that a narrow content must satisfy in order to be considered adequate. I then choose my favoured construal and give a version of it which will hopefully be seen as an improvement on the other construals (by satisfying all the adequacy conditions). It will be argued that my version of the narrow content construal will provide, not just an adequate narrow content, at least in relation to a subject's perceptual experience, but perhaps also necessary and sufficient conditions for a mental state, such as a perceptual belief, to have the specific content that it has.

## ACKNOWLEDGEMENTS

Firstly I would like to thank my supervisor Professor Steve Edwards for his invaluable help and guidance over the last five years. Without Professor Edwards I would not have been able to even attempt a PhD, as he kindly agreed to be my supervisor when it seemed that my plans for future study were fizzling-out for various reasons. I will always be grateful for the kindness and generosity he has shown me over these years. I would also like to thank Dr Mike McNamee for all his kindness and support.

I am also very grateful to the School of Health Science for taking me on as a post-graduate student which allowed me to study there under Professor Edwards, and for providing me with a bursary for the five years, as I would not have been able to afford it otherwise.

I would like to thank everyone who attended a small seminar for Philosophy and Health Care at Swansea University, where I read an excerpt from my work and gained much useful feedback from those present. Also many thanks to Dr Alessandra Tanesini for inviting me to give a talk in the Philosophy Dept. of Cardiff University. I was very nervous but everyone was very kind to me and the issues raised by those present were very useful and gave me much to think about.

I would like to thank my colleagues and friends in the Library and Information Centre for all their support, especially those at the Issue Desk and in Inter-library Loans.

Finally, I would like to thank my family, Mum and Dad, sister Karen and niece Natalie, and Gabby (the dog) for all their love and support.

P.S. Many thanks to Maggie at Gnomeland.co.uk for the garden-gnomes.

# CONTENTS

INTRODUCTION	1
PART 1 CONCEPTUAL ROLE NARROW CONTENT	9
LOAR'S PSYCHOLOGICAL CONTENT	10
1.1 <b>The Conception of Psychological Content</b>	10
1.1.1 <i>Burge's Objections to Conceptual Role Individuation of Thoughts</i>	12
1.1.2 <i>The Relationship Between De Dicto Belief Ascriptions and Psychological Content</i>	16
1.1.3 <i>Loar's Response to Burge's Two Objections From the "Arthritis" Example</i>	18
Loar's Response to Objection 1	20
Loar's Response to Objection 2	23
1.2 <b>Assessments and Criticisms of Loar's Narrow Psychological Content</b>	26
1.2.1 <i>Fregean Puzzles, Interpersonal Beliefs and Psychological Generalisations</i>	26
1.2.2 <i>Assessment of Loar's Use of Pierre-Type Examples</i>	32
What Conclusions Can be Drawn From Loar's Use of the Pierre-Type Examples?	33
A Difference of Concepts, But a Type-Identity of Beliefs	37
Notes	49
AN EVALUATION OF LOAR'S PSYCHOLOGICAL CONTENT IN RELATION TO BURGE'S "ARTHRITIS" THOUGHT EXPERIMENT	53
2.1 <b>An Attempt to Specify the Narrow Psychological Content of the Belief Ascriptions in Burge's "Arthritis" Thought Experiment</b>	59
2.1.1 <i>The Actual Situation</i>	59
Burge's De Dicto/De Re Distinction	61
The Possibility of Subjective and Objective De Re Beliefs?	63
The Possibility of Subjective and Objective De Dicto Beliefs?	67
2.1.2 <i>The Counterfactual Situation</i>	77
A Possibility of Narrow Belief Content?	79
Notes	84

LOAR, NARROW PSYCHOLOGICAL CONTENT AND INTENTIONALITY	87
3.1 <b>Loar's Journey to his "Context-Indeterminate Realization Conditions"</b>	89
3.1.1 <i>The Realization Conditions of a Belief</i>	91
3.1.2 <i>The Context-Indeterminate Truth Conditions of a Belief</i>	93
3.1.3 <i>The Context-Indeterminate Realization Conditions of a Belief</i>	95
3.2 <b>Assessment and Criticisms of Loar's Realization Conditions</b>	98
3.2.1 <i>An Explication of Realization Conditions in Terms of Possible World Semantics</i>	98
3.2.2 <i>Stalnaker's Critique of Loar's Realization Conditions</i>	104
3.3 <b>PART 1 CONCLUSION</b>	118
<b>Notes</b>	122

**PART 2 INDEXICAL NARROW CONTENT 124**

FODOR'S "MAPPING THEORY"	125
4.1 <b>Methodological Individualism and Causal Powers</b>	125
4.1.1 <i>Possible Problems for Fodor's Cross-Context Test</i>	127
4.2 <b>Does Identity of Thought-Content Determine Identity of Extension?</b>	130
4.2.1 <i>How Fodor Repairs the Connection Between the Content Identity and Extensional Identity of Thoughts</i>	132
4.3 <b>The Extensional Identity Criterion for Thought-Contents</b>	136
4.3.1 <i>The Extensional Identity Criterion for Narrow Content</i>	137
4.4 <b>How Does Fodor Specify his Narrow Content?</b>	139
<b>Notes</b>	144

ASSESSMENT AND CRITICISMS OF FODOR'S NOTION OF NARROW CONTENT	147
5.1 <b>What is a Function? – (i)</b>	147
5.2 <b>Does Fodor's Narrow Content Collapse Into Syntax?</b>	152
5.2.1 <i>What is a Function? – (ii)</i>	158
5.3 <b>Objections to the Modified Construal of Fodor's Mapping Theory</b>	169
5.3.1 <i>Contexts of Acquisition and the Causal-Historical Theory of Reference</i>	170
5.3.2 <i>Possible Problems Relating to My Construal of an Intensional-Function</i>	180
Wide Conceptions	181
A Conception of Narrow Content That Depends on Another Conception of Narrow Content	184
5.4 <b>PART 2 CONCLUSION</b>	186
<b>Notes</b>	190

**RT 3      NARROW REPRESENTATIONAL CONTENT      192**

DENNETT'S NOTIONAL WORLDS	193
6.1 <b>How Dennett Proposes to Formulate a Subject's Notional World</b>	196
6.2 <b>Assessments and Criticisms of Dennett's Notional Worlds</b>	199
6.2.1 <i>Loar's Views on Notional Worlds</i>	199
6.2.2 <i>A Clarification of Dennett's Notional Worlds Conception</i>	203
6.2.3 <i>Notional Worlds and Representational States</i>	207
<b>Notes</b>	210

A NARROW REPRESENTATIONAL REFORMULATION OF DENNETT'S NOTIONAL WORLDS	212
7.1 <b>A Far-Fetched Thought Experiment</b>	212
7.2 <b>A Cunning Plan for the Objective Determination of a Notional World</b>	216
7.3 <b>Notional World Formulation Using the Cunning Plan</b>	219
7.4 <b>Specifying the Organismic (Narrow Content) Contribution to a Person's Propositional Attitudes</b>	226
7.4.1 <i>Can Notional Beliefs be Given a De Re Specification?</i>	228
7.4.2 <i>Can Notional Beliefs be Given a De Dicto Specification?</i>	229
Objections to the Stereotype-Strategy of Specifying Narrow Content	236
7.5 <b>PART 3 CONCLUSION</b>	242
<b>Notes</b>	245

**RT 4      A PROPOSAL FOR AN ADEQUATE AND NECESSARY NARROW CONTENT      249**

THE FORMULATION OF AN ADEQUATE NARROW CONTENT	251
8.1 <b>Narrow Content as a Subjective Virtual Reality</b>	251
8.2 <b>The Semantic Properties of the Narrow Representational Content</b>	255
8.2.1 <i>Narrow Meaning/Content Properties</i>	256
8.2.2 <i>Narrow Truth Conditional Properties</i>	262
An Important Objection to the Above Construal of Narrow Truth Conditions	265
A Relevant Example From The Literature	270
8.2.3 <i>Narrow Referential Properties</i>	280
Narrow Content and Direct Reference	282
Russell's Notion of Acquaintance	292
Acquaintance and the Possibility of "Direct" Reference?	297
8.3 <b>Narrow Content and Psychological Explanation</b>	299
<b>Notes</b>	304

THE POSSIBILITY OF NECESSARY AND SUFFICIENT NARROW CONDITIONS FOR BELIEF CONTENT	310
9.1 <b>Necessary and Sufficient Conditions of Narrow Content</b>	311
9.1.1 <i>A Necessary Condition</i>	314
9.1.2 <i>A Sufficient Condition</i>	317
9.2 <b>A First Possible Objection to the Above Arguments for Necessary and Sufficient Narrow Content</b>	319
9.2.1 <i>McGinn's "Percy" Thought Experiment</i>	320
9.2.2 <i>Davies' Variation of McGinn's "Percy" Thought Experiment</i>	323
9.2.3 <i>How Davies' Variation of McGinn's Thought Experiment     Could be Construed as an Objection to My Arguments</i>	328
9.2.4 <i>McGinn's "Weak Externalism" and Narrow Representational     Content</i>	330
9.3 <b>A Second Possible Objection to the Above Arguments for Necessary and Sufficient Narrow Content</b>	333
9.3.1 <i>The Possibility of Absent and Inverted Qualia</i>	336
9.3.2 <i>How the Consequences of Wide Externalist Representationalism     Impact on My Positions in Situation (i) and for Necessary/Sufficient     Narrow Content</i>	343
Situation (i)	343
Necessary/Sufficient Narrow Content	347
9.4 <b>Narrow Representational Content and Phenomenal Intentionality</b>	352
9.5 <b>Necessary/Sufficient Narrow Content and Explanations of Intentional Behaviour</b>	362
Notes	368
MAIN CONCLUSION	371
BIBLIOGRAPHY	380

Phenomenalism would be a way – the only way – to avoid the moral of the Twin Earth arguments; we can avoid an externalist account of content only by moving the external world itself into the mind.

- Robert C. Stalnaker, from “Twin Earth Revisited.” *In Context And Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press, p.217.

Designers of virtual reality machines trade on the fact that the brain cannot, as we might put it, see beyond the local nature of its stimuli to the nature of their distal origins. These designers set themselves the task of producing from their machines stimuli at our peripheries exactly like those that would be produced by various external environments – the Sahara, perhaps – and to the extent they succeed, we have an experience as of being in the Sahara.

- David Braddon-Mitchell / Frank Jackson, from *The Philosophy of Mind and Cognition*. Oxford; Cambridge, Massachusetts: Blackwell Publishers, p.212.

NEO: This – This isn't real?

MORPHEUS: What is real? How do you define real? If you're talking about what you feel, taste, smell, or see, then real is simply electrical signals interpreted by your brain.

- Part of an exchange between Neo and Morpheus from the film *The Matrix* (1999).

# INTRODUCTION

This thesis will examine the notion of narrow content, as it has been conceived of mainly in the philosophy of mind. Narrow content can be defined as the following:

Narrow content is a kind of mental state that is shared between doppelgangers or two internal physical duplicates in Twin Earth-type situations.

In talking of “Twin Earth-type” situations, these are situations similar to Putnam’s (1975) original Twin Earth thought experiment.<sup>1</sup> The gist of the thought experiments is that two physical duplicates are located in different environments, which appear identical to the duplicates, but which differ in respect of underlying physical structure or social/linguistic conventions. For example, in Putnam’s original thought experiment, the underlying physical difference involves water, where the duplicates’ physical environments both contain liquids called water which look, feel, taste, etc. identical, but have different chemical compositions, on Earth, water is chemically composed of H<sub>2</sub>O, whilst on Twin Earth, water is chemically composed of XYZ.

An example of a thought experiment involving differences in the social environment is given by Burge (1979), where it is imagined that in one situation the term “arthritis” refers to a rheumatoid ailment of the joints, whilst in another situation the term “arthritis” refers to a rheumatoid ailment of the joints and bones in general.

Moreover, in the two kinds of thought experiments the different physical or social



environments could be described in terms of being two different planets in the same universe (as in Putnam's original thought experiment) or they could be understood in counterfactual terms, where the first situation is considered to be the actual Earth, and the second situation is considered to be a counterfactual Earth.

One conclusion drawn from both Putnam's and Burge's thought experiments is that the contents of at least some thoughts, beliefs, etc. are not determined solely by the subject's internal physical properties, but are also determined by the subject's environment, either physical or social. Putnam argued that because of the difference in the chemical compositions of the Earth water and Twin Earth water, the meanings of the "water" terms are also different. Identical utterances of, say, the sentence "Water is wet" on Earth and Twin Earth would have different meanings, i.e., on Earth the utterance would mean that "H<sub>2</sub>O is wet", whilst on Twin Earth the utterance would mean that "XYZ is wet". One can also see how this conclusion could be extended to the beliefs that are being expressed by the subjects' utterances, so the beliefs could be considered to have different content.<sup>2</sup> However, it was Burge who made this last point explicit with his "arthritis" thought experiment, where the difference of *arthritis* concepts intuitively leads to differences in the duplicates' propositional attitudes containing the term "arthritis".

The upshot of both Putnam's and Burge's thought experiments is that it seemed that internalism with regards to mental content had been refuted, where internalism is understood as the view that the contents of mental states are determined only by the properties internal to the subject.<sup>3</sup> As the subjects concerned in these thought experiments are internal duplicates, then if internalism is true, it would be expected

that the contents of their mental states would also be the same, despite the differences in their environments. However, the strong intuition fuelled by Putnam and Burge is that at least some of the mental contents of the subjects would be different, thus leading to the apparent refutation of internalism. This led to the popularity of externalism with regards to mental content, which is the view that the identity of the contents of at least some mental states is determined by the subject's physical and/or social environment.

However, due to the plausibility of the idea of the local supervenience of the mental on the physical, there is also a strong intuition that in both thought experiments, there is a type of mental content that is shared between the duplicates and which is determined only by the identical internal physical properties of the duplicates. It is this mental content which is called narrow content, as opposed to the mental content which is considered different between the duplicates, which can be called wide content.<sup>4</sup> Therefore, throughout this thesis we will be concerned with examining exactly what type of mental content, if any, remains the same in the duplicates in Twin Earth-type situations, due to the identity of their internal physical properties. Moreover, because of this identity of internal physical properties, one can make the following necessity claim about narrow content: if a subject A instantiates a narrow content NC, then an internal physical duplicate of A will necessarily also instantiate that same narrow content NC.

A related point to this identity of narrow content is that it should also result in an identical individuation of narrow psychological states which adequately explain the consequent behaviours of the duplicates. Therefore, in Twin Earth-type situations,

there should be a narrow psychological explanation, which only invokes narrowly individuated mental states, and which subsumes the behaviours of the duplicates.

However, as well as the motivation for narrow content that comes from Twin Earth cases, another motivation comes from so-called Frege cases and thought experiments which are influenced by Frege, such as Kripke's (1979) "Pierre" example. The basic gist of the aforementioned is that, for example, a subject refers to the same object in the world by two different referring expressions, but does not realise that the expressions refer to the same object and so believes that there are two distinct objects in the world. For instance, using Frege's (1892) original example, someone might believe that the sentence "The morning Star is a body illuminated by the Sun" is true, whilst believing that the sentence "Venus is a body illuminated by the Sun" is false, even though the morning star is the planet Venus. Frege's point is that as "morning star" and "Venus" are coreferential expressions, if the meaning of the beliefs expressed by the aforementioned sentences is completely determined by their reference, then the meaning of the beliefs should be the same. However, this does not seem to be the case, as it seems plausible that someone could believe that one sentence is true whilst the other is false, and not be contradicting himself or herself.

For Frege, this meant that whilst the meaning of a thought might well be partly determined by its "reference" to things in the world, the meaning could also be partly determined by the "sense" or mode of presentation of the object involved with the thought. With the example of the sentences containing the expressions "morning star" and "Venus", it would be said that the beliefs expressed by both sentences had the same reference, but different senses. The relevance that this has for positing some

sort of narrow content is that the individuation of one's beliefs is not completely determined by objects in the external world, but also depends, to some degree, on the mode of presentation that the person has.<sup>5</sup> Putting it another way, belief-contents are also individuated according to how that person conceives or represents the objects in the world which the beliefs are about.

Therefore, in order to examine and specify the notion of narrow content, this thesis will have the following structure. The first three parts examine three of the most popular construals of narrow content in the philosophy of mind, namely, conceptual role, indexical and representational. In each part an exemplar of the narrow content construal will be chosen and examined in detail. In the fourth part, I give my own preferred construal of narrow content which seems the most plausible to me, and which will (hopefully) be considered an improvement on the three previous construals. To help achieve this aim, I will employ the following three *conditions of adequacy* that a construal of narrow content has to satisfy in order to count as an adequate narrow content: (1) The narrow content construal must make it plausible that there is such a thing as an internal, "in the head", type of mental content, that is shared by duplicates in Twin Earth-type thought experiments. (2) The narrow content construal must be such that a plausible specification of that content can be given. (3) The narrow content construal must provide an adequate psychological generalisation that explains the behaviour of both duplicates. Additionally, the construal must be flexible enough to make more fine-grained distinctions in mental content, e.g., when dealing with Frege cases.

What we will find as the thesis progresses is that none of the construals of narrow content is pure; they sometimes include aspects from the other construals. For instance, the indexical construal contains elements from the conceptual role construal, the representational construal contains elements from the indexical construal, etc. Nevertheless, to give the thesis some sort of structure and order, I decided to categorise the types of narrow content as I have done, as this seemed the most straightforward way to proceed.

The first kind of narrow content to be examined is a conceptual role one. This type of narrow content individuates beliefs by the conceptual role the belief plays in the subject's mental framework, i.e., the way that a belief inferentially relates to sensory inputs, other beliefs, and the consequent behaviour of the subject. The exemplar of a conceptual role narrow content will be Loar's (1988a, 1988b) notion of the psychological content of a belief, which he argues is distinct from the content a belief has on a de dicto reading.<sup>6</sup>

The second kind of narrow content to be examined is an indexical one. This construal usually equates narrow content with Kaplan's (1989) notion of the character of an indexical expression. For Kaplan, the character of an indexical expression is a function from contexts of utterance of the expression to the content the expression has. The exemplar of an indexical narrow content will be Fodor's (1987) "mapping theory", where the narrow content of a thought is considered to be a function from contexts of thought onto wide content.

The third kind of narrow content to be examined is a representational one, where the narrow content of a belief is considered to originate from the way the subject represents things as being, or of how things seem to him/her.<sup>7</sup> The narrow representation of how things seem to the subject usually involves the perceptual experience of the subject, e.g. visual perception. The exemplar of a representational narrow content will be Dennett's (1982) idea of the notional world of a subject.

In part four, I put forward my own preferred construal of narrow content, which is mainly influenced by the representational narrow content which comes from a variation of Dennett's conception. The narrow content of a subject's belief originates from the internal representations that are produced from the proximal stimulations on the sensory receptors of the subject. I argue that from the first-person perspective of the subject, these internal representations form a narrow virtual reality of the distal state of affairs before the subject.

## Notes to Introduction

<sup>1</sup> The notion of narrow content first came to prominence in Putnam (1975), where a narrow and wide bifurcation of the content of a person's psychological states was suggested. For Putnam, a narrow psychological state was one that involved the assumption of "methodological solipsism" (MS), where the narrow state in question did not presuppose "...the existence of any individual other than the subject to whom that state is ascribed" (1975, p.220). The upshot of MS seemed to be that if a subject had a narrow psychological state concerning a certain physical object or event, this did not logically entail that the object or event actually existed in the external world.

<sup>2</sup> In this thesis, when I refer to the "content" or "meaning" of a belief, I will take the two terms to be synonymous.

<sup>3</sup> As far as I can discern from the literature, the internalist view on the determination of, say, a belief content, is the following. For a subject to have a belief with a certain content, this can involve causal-historical relations between the subject and the external world, but how the content is determined or constituted is a matter of how the subject is internally. That is, the determination of the belief-content is not considered to be logically dependent on the subject's immediate environment. The content of a subject's belief could stay the same even though his/her immediate environment might change radically. It is this definition of internalist belief-content determination that I will be utilising in my later arguments for a plausible narrow belief-content. See Braddon-Mitchell/Jackson (1996) p.212 for an explication of this internalist position with regards to belief-content. For a statement of this position in relation to a subject's qualitative experience, see Rey (1998) p.439 and Levine (2003) p.61.

<sup>4</sup> The wide content of our beliefs (and other propositional attitudes) is usually called referential or truth conditional content. Wide belief-content is referential as it refers to objects and states of affairs in the world; and it is also truth conditional content because of this fact. That is, the truth or falsity of a belief

depends on whether the wide content of the belief is true or false, where the latter depends on whether the state of affairs in the world that the belief is about does actually obtain or not. For example, the truth conditions of my belief that water is wet would be that H<sub>2</sub>O is wet, whereas the truth conditions of my duplicate's belief that water is wet would be that XYZ is wet. Therefore it can be seen that the truth conditions of our water-beliefs are related to the wide content of our water-beliefs, with the result that because the water-beliefs have distinct wide content, they also have distinct truth conditions. At least, this is how the story goes. As this thesis progresses, we shall discover whether narrow content can also be referential and truth conditional content.

<sup>5</sup> It might be claimed here that Frege intended his senses to be "objective" in that they include components that necessarily make reference to objects and states of affairs in the external world. The upshot of this is that Fregean senses could be considered externalist or object-dependent in nature. However, I take the view that one could have a Fregean sense with external world components that did not logically imply that the object it referred to existed in the external world, e.g., one could have a Fregean sense about a mythical creature like, say, a "liger", which was a magical creature considered to be a cross between a lion and a tiger, belief in which was popular in Napoleonic times. It is in this sense that I think Fregean senses could support the notion of narrow content, even though they may consist of "objective" components.

<sup>6</sup> A *de dicto* belief is considered to be a belief that a certain *dictum* or proposition is true. The propositional content of a *de dicto* belief can be individuated not only by the external world object or state of affairs that the belief is about, but also by the way that the subject conceives of that object or state of affairs. It is because of this that a *de dicto* reading of a belief is said to characterise the belief content for the subject. In contrast, a *de re* belief is considered to be a belief about a particular object or state of affairs, that it has a particular property. A *de re* belief is only individuated by the external world object or state of affairs that the belief is about, with the subject's conceptions possibly not playing any individuating role. For example, consider the belief that the Evening Star is a body illuminated by the sun and the belief that the Morning Star is a body illuminated by the sun. On a *de dicto* reading, these beliefs could be considered type-distinct, particularly if the person having them did not know that the Evening Star and Morning Star are the same object, namely, the planet Venus. That is, the beliefs could be individuated differently according to the person's different conceptions of Venus. However, on a *de re* reading, the beliefs would be considered type-identical, as they would be about the same object, Venus. Hence, the different conceptions of Venus would play no individuating role on a *de re* reading of the beliefs.

<sup>7</sup> In this thesis, the "representational content" and "perceptual content" of a mental state will be considered as synonymous. The perceptual content of a mental state will be how the state represents things as being in the subject's environment (see Edwards, 1994, p.96).

# **PART 1**

## **CONCEPTUAL ROLE NARROW CONTENT**

In the first part of the thesis I will examine a conceptual role construal of narrow content, the exemplar of which is Loar's notion of psychological content. In chapter 1 I will explicate Loar's psychological content and how it deals with strong objections to the conceptual role individuation of belief content given by Burge. I then consider more specific criticisms of Loar's psychological content that deals with the methods he uses to support his arguments for it. In chapter 2 an evaluation of Loar's psychological content is made, and I raise what I see as a problem for Loar's notion, namely, that it cannot be given a plausible specification. I then attempt to provide a way of specifying the psychological content whilst applying Loar's notion directly to Burge's "arthritis" thought experiment. In chapter 3 Loar's attempt to provide his narrow psychological content with intentionality is explicated and assessed.



## CHAPTER 1

### LOAR'S PSYCHOLOGICAL CONTENT

#### 1.1 The Conception of Psychological Content

Loar (1988a, 1988b) doesn't actually call his conception "narrow content", but labels it "psychological content". However, it seems fairly clear that the latter content is a type of narrow content, as Loar views it as "content" which is shared between doppelgangers or the same person in a counterfactual situation, in the standard Putnamian and Burgean thought experiments. According to Loar, "psychological content" means:

...whatever individuates beliefs and other propositional attitudes in commonsense psychological explanation, so that they explanatorily interact with each other and with factors such as perception in familiar ways (Loar, 1988a, p.99).

By commonsense psychological explanation, Loar just means the ordinary way that the interaction of various beliefs and desires explains why people behave as they do. For example, if someone desired  $x$ , and believed that in order to bring about  $x$  he/she would have to do  $y$ , then this would *explain* that person's doing  $y$ . Nevertheless, from the above quote, it seems clear that Loar's "psychological content" will be individuating beliefs, etc., according to the conceptual role they play in the mental framework of the subject. This is due to the fact that conceptual role semantics (CRS) is a way of individuating thoughts (and other psychological states like beliefs, desires,

etc.) by the inferential relations that those thoughts have with other thoughts of the subject, and with the sensory inputs and behavioural outputs of the subject.<sup>1</sup>

Moreover, Loar declares that his thesis of psychological content is generally understood to involve the individuation of a person's propositional attitudes "...along non-social, individualistic lines" (Loar, 1988a, p.100).

Now, it might be questioned whether Loar's psychological content *could* have this independence from the social environment. It seems reasonable to say that "commonsense psychological explanation" might be independent from the physical environment, but surely the commonsense individuation of beliefs would be dependent on their standard roles in the subject's linguistic community. I think that what Loar means is connected with the view he takes of how we explain and predict other peoples' behaviour. What he calls "commonsense psychological explanation" concerns how beliefs interact with each other, from the subject's point of view, and which result in certain behaviour by the subject. According to Loar, we understand other peoples' behaviour by "projecting our self-understanding" onto them (Loar, 1988b, p.128). A similar way of putting it would be that we try to adopt another person's subjective point of view, to try to imagine "what it would be like" for us from their viewpoint.<sup>2</sup> So, Loar's notion of "commonsense" in relation to psychological explanation, does seem to revolve around the person's conception of his/her own thoughts, and how this subjective perspective results in various structured interactions among the person's thoughts to produce behaviour.

Of course, it could still be claimed that what is usually meant by "commonsense" psychological explanation is not what Loar says it means, but that it normally

involves that-clause ascriptions and the social environment. In that case, perhaps Loar should have just said that psychological content played a role in “subjective” or “individualist” psychological explanations. Be this as it may, the important point is that for Loar, it is the subject’s conception of things which individuates his/her thoughts, and this conceptual role-individuation may differ from the pattern of thought-individuation that would result from a social/linguistic viewpoint.

### 1.1.1 *Burge’s Objections to Conceptual Role Individuation of Thoughts*

However, as Loar acknowledges, the idea of individuating propositional attitudes by their conceptual roles has faced very influential criticism from Burge (1979). It is in examining Burge’s arguments against the conceptual role individuation of propositional attitudes, that Loar explicates his own thesis of “psychological content”. The strategy used to criticise conceptual role individuation of thoughts is usually to show, by the use of a particular example, how this type of individuation is at odds with how we would normally ascribe thoughts using de dicto or oblique that-clauses (Loar, 1988a, p.99). Loar identifies two objections of this type contained in Burge’s well-known “arthritis” thought experiment:

*Objection 1:* that two beliefs which are individuated differently according to conceptual role theory, count as the same beliefs according to commonsense ascription. (Put more formally, identity of conceptual role is not necessary for identity of belief ascriptions).

This objection comes about in the following way. It will be remembered that in Burge's "arthritis" thought experiment, a person, call him Bert, believes that he has arthritis in various parts of his body. He believes that he has arthritis in his ankles, but also, importantly, he believes that he has arthritis in his thigh. Bert is unaware that arthritis is defined as an ailment that can only affect joints, and believes, mistakenly, that it is an ailment of bones in general. It is only when a doctor explains this to Bert, that he realises that while his other arthritis-beliefs could be considered true, his belief that he had arthritis in the thigh was a false belief.

The question now is whether Bert's earlier belief that he had arthritis in his ankles (before the doctor corrected him about his use of the term "arthritis") is to be considered as the same belief which the doctor has about Bert, i.e., that he has arthritis in his ankles. Now, Burge (according to Loar) takes the view that if the beliefs are individuated according to conceptual role, then they would be viewed as different beliefs (Loar, 1988a, p.100). This is because the concept of "arthritis" that Bert uses, and the way it interacts with other beliefs and thoughts, which might produce certain behaviour, differs from the concept of "arthritis" used by the doctor, and the corresponding way it interacts with the doctor's other beliefs and thoughts, etc.

However, despite the two beliefs being distinct by conceptual role-individuation, Burge could quite intuitively claim that commonsense attitude ascription using oblique that-clauses would count Bert's belief and the doctor's belief as the same, i.e., he (Bert) believes that he has arthritis in his ankles. According to Loar, this point means that "...sameness of conceptual role is not *necessary* for sameness of psychological content" (Loar, 1988a, p.100). As the conceptual roles in this case are

considered different, Loar attributes to Burge the view that belief ascriptions using oblique that-clauses are what characterise the psychological content of a belief, i.e., two beliefs which have the same oblique that-clause ascriptions, also have the same psychological content. It should also be noted here that according to the definition that Loar gives to “psychological content”, the above conclusion means that it is being claimed that the content of oblique that-clauses is what individuates beliefs in commonsense psychological explanation.

*Objection 2:* that two beliefs which are individuated as the same according to conceptual role theory, count as different beliefs according to commonsense ascription. (Put more formally, identity of conceptual role is not sufficient for identity of belief ascriptions).

This objection comes about in the following way. Burge now imagines a counterfactual world, where the definition of “arthritis” is that it is an ailment of bones in general, and not just of joints. We might call this counterfactual version of arthritis by a different name, e.g., “tharthritis”. In this counterfactual situation is placed the long-suffering Bert, who is considered to be individualistically identical to the Bert of the previous actual situation. That is, counterfactual-Bert is not just neurophysiologically identical, but his mental states and the conceptual roles that they play in his mental life, are also identical. Now, in this situation, if we individuate beliefs according to conceptual role, then Bert’s actual beliefs would have to be the same as his counterfactual beliefs, because the conceptual roles are identical. This means that in the actual and counterfactual situations, according to conceptual role-individuation, Bert would have the same belief, i.e., “I have arthritis in my thigh”.

However, in contrast, Burge is now able to claim that commonsense ascription would count Bert's actual and counterfactual beliefs as different. In the actual situation, the oblique that-clause ascribing Bert's belief would be something like "He believes that he has arthritis in his thigh"; whereas, in the counterfactual situation, the oblique that-clause ascribing Bert's belief would be something like "He believes that he has tharthritis in his thigh". So, with the individualistic conceptual roles of Bert's two beliefs being the same, and yet, according to commonsense belief ascription, the psychological content of the two beliefs seeming to be different, this enables the claim to be made that "...sameness of conceptual role is not *sufficient* for sameness of content..." (Loar, 1988a, p.100).

A number of points need to be noted here. Firstly, as in Objection 1, Burge is (according to Loar) equating the psychological content of a belief, i.e., the content that interacts with other beliefs in commonsense psychological explanation, with the content that an oblique that-clause ascription of the belief contains. And secondly, that the apparent difference in the content of the two beliefs, seems to be due, not to individualistic differences, but to conceptual differences in the social environment of the subject. Of course, it is interesting to note that the type of "conceptual" difference that I just mentioned, can only be a *social-conceptual* difference, as the *individualist* concepts of Bert were kept the same.<sup>3</sup> This is a clue as to the strategy that Loar is going to adopt in response to these objections against the individuation of thoughts and propositional attitudes by their conceptual roles.<sup>4</sup>

### 1.1.2 *The Relationship Between De Dicto Belief Ascriptions and Psychological Content*

Loar's analysis of Burge's criticism of the conceptual role individuation of beliefs in his "arthritis" thought experiment leads him to believe that:

The question then to be addressed is the relation between *de dicto* or oblique ascriptions of beliefs and their psychological contents, between such ascriptions and their individuation in commonsense psychological explanation. I shall argue that psychological content is not in general identical with what is captured by oblique that-clauses, that commonsense constraints on individuation induce only a loose fit between contents and that-clauses... (Loar, 1988a, pp.101-102).

It is important to remember that in the above quotation from Loar, when he uses the term "commonsense", I don't think he necessarily means to include any factors from the social environment. Arguably, Loar takes commonsense psychological explanation to consist in the structured relations between a subject's beliefs and other attitudes, and how these relations result in certain behaviour. In Loar's opinion, the social factors from the subject's environment are usually contained in the content of the oblique that-clause ascriptions. However, Loar begins addressing the above question by identifying what seems to be an explicit assumption in the arguments of Burge (and Stich); that is, if two beliefs have the same *de dicto* ascriptions, then they will also have the same psychological content (Loar, 1988a, p.102). Loar refines the explicit assumption to the following:

(A) Sameness of the de dicto or oblique occurrence of a general term [or proper name] in two belief ascriptions implies, if everything else is the same, sameness of the psychological content of the two beliefs thus ascribed (Loar, 1988a, p.102; Loar, 1987, p.92).<sup>5</sup>

He also identifies the related assumption which is the converse of (A), namely:

(B) Differences in de dicto or oblique ascription imply differences in psychological content (Loar, 1988a, p.102; Loar, 1987, p.92).

In Burge's "arthritis" thought experiment, it was seen that Objection 1 claimed that while individuation of two beliefs by their conceptual roles would result in the beliefs being judged as distinct, commonsense ascription seemed to individuate the beliefs as the same, i.e., "Bert believes that he has arthritis in his ankles" and "The doctor believes that Bert has arthritis in his ankles". So here it appears that Loar is attributing (A) to Burge's reasoning, with the general term "arthritis" having the same oblique that-clause ascription in both cases, resulting in the beliefs having the same "psychological content". Similarly, with Objection 2 of Burge's argument, it can be seen that (B) seems to come into play; that is, because the oblique that-clause ascriptions of the two beliefs are different (one that-clause ascribes to Bert the belief that he has arthritis, while the other that-clause ascribes to Bert the belief that he has tharthritis), then the psychological content of the beliefs are also different. In each case, Burge does not seem to associate the function of psychological explanation to the conceptual roles that the beliefs have in the subjective understanding of the person whose beliefs they are.<sup>6</sup>



### 1.1.3 *Loar's Response to Burge's Two Objections From the "Arthritis" Example*

In order to respond to Burge's criticisms, Loar's strategy is to first introduce two thought experiments, the first dealing with proper names and the second dealing with general terms, which he hopes will show that we must reject (A). The thought experiment dealing with proper names is a variant of Kripke's (1979) famous "Pierre" example. We are asked to imagine someone called Pierre, who lives in France, and who learns of a place called "Londres", from which time he comes to believe and assert that "Londres est jolie". Some time later, Pierre pays a visit to London, not realising that "London" is the same place as the "Londres" which he has learnt about. His visit is a short one, and he only sees a rather shabby and squalid part of the city during this time. Because of this, he comes to believe and assert "London is not pretty". Now, on an oblique univocal reading, it seems that we can ordinarily ascribe two contradictory beliefs to Pierre, namely, that he believes London is pretty, and that he believes London is not pretty (Loar, 1988a, p.102).<sup>7</sup>

Loar's variation on this example is now to imagine that when Pierre pays his visit to London (still not knowing that it is the same place as Londres), he doesn't go to a shabby part of the city, but instead goes to a more picturesque part, say, Canary Wharf. This leads him to believe and assert that "London is pretty". Now, in this case, on an oblique univocal reading, it seems that we must say that only one belief can be ascribed to Pierre, i.e., that he believes London is pretty.

However, Loar is of the opinion that if we individuate beliefs according to commonsense psychological explanation, then what we have are two distinct beliefs,

as Pierre's beliefs about Londres and London will potentially interact with other beliefs in different ways. For example, Pierre might believe that Buckingham Palace is in Londres, but not in London, and so on. It can be seen that this conclusion is contra to (A), i.e., despite the same oblique belief descriptions ("Londres is pretty" and "London is pretty"), the two beliefs seem to have different psychological contents. And Loar believes that this individuation of the two beliefs not only means that their psychological or conceptual roles are different, but that there is also a real difference of "content" between them. He attributes this difference to the way that Pierre conceives of things, how he takes things to be in the world, from his subjective point of view (Loar, 1988a, p.103). It is also possible that this could be seen as a difference in intentional content, as Pierre's beliefs seem to be "about" objects and states of affairs in the world.

Loar's thought experiment involving general terms is similar to the argument involving proper names, in that it again involves the same object or thing being known by a term, which, due to linguistic differences, is thought to apply to two distinct objects instead of the one. The difference is of course that as it is now a general term that is being dealt with, it is a certain class of objects which the term is being used to refer to. The upshot of the thought experiment is the following. Paul, who speaks English, comes to believe that the class of animals that we would call "cats", is actually called "chats" (which is the French word for "cats"). Apart from this point, Paul has many other correct beliefs about cats, and has what Loar describes as "...a perfectly good recognitional acquaintance with cats" (Loar, 1988a, p.103). He might even come to believe and assert that "All chats have tails". However, Paul also comes to learn things about a class of animals which he calls "cats", not realising

that this latter class of animals is the same as the previous class which he calls “chats”. Similarly, Paul comes to believe and assert that “All cats have tails”.

Thus, it could be claimed that Paul’s two beliefs that all chats have tails and that all cats have tails, would, if individuated on an oblique univocal reading, count as the same belief, i.e., all cats have tails. However, according to Loar, if we individuate beliefs according to their conceptual role, then this will result in the two beliefs being considered as distinct. Also, Loar believes that this conceptual role-individuation of the beliefs is more in keeping with the workings of commonsense psychological explanation, as the two beliefs will interact differently with other beliefs that Paul has, and so have distinct psychological contents (Loar, 1988a, p.103). For example, Paul might believe that cats like chasing mice but that chats do not, and so on. Moreover, Loar believes that the distinct conceptual roles of Paul’s two beliefs also intuitively signal a distinctness of representational content between the beliefs. As Paul conceives of things, his two beliefs represent the world as containing “two distinct facts”, i.e., facts concerning cats and chats (Loar, 1987, p.93).<sup>8</sup>

After Loar has pumped the intuition that (A) could be false with the above thought experiments, he then turns his attention specifically to the two objections against conceptual role individuation of beliefs that were raised by Burge’s “arthritis” thought experiment.

Loar’s Response to Objection 1

Objection 1 was essentially this: where individuation by conceptual roles would result in two particular beliefs being counted as distinct, individuation by “commonsense” ascription results in the same two beliefs being counted as the same; the reason for this latter conclusion being that both beliefs had the same oblique, univocal, that-clause ascriptions. By this, Loar took Burge to mean that the two beliefs had the same psychological content, despite their distinct conceptual roles, i.e., the beliefs would be considered the same as far as commonsense psychological explanation was concerned. The upshot of all this was that: *For Burge the content of a belief that played the role in psychological explanation, was contained in its oblique that-clause ascription.* From the earlier “arthritis” example, the two beliefs in question were Bert’s and the doctor’s, and the same oblique that-clause ascriptions they had were “believes that Bert has arthritis in his ankles”.

In answering Objection 1, Loar uses another “Paul” example, which again invokes the use of an English word and French word which both refer to the same concept, being mistakenly thought to refer to distinct concepts. To summarise the situation, Paul believes that he has the rheumatoid ailment called “arthrite” in his thigh and ankles, and, because of hypochondriac tendencies (in addition to his linguistic confusions), that he also has “arthritis” in his ankles (Loar, 1988a, p.104). According to Loar, the result of this situation is the following: even though on an oblique univocal reading, Paul’s two beliefs would be counted as the same, i.e., “believes that he has arthritis in his ankles”; the two beliefs will not be counted as the same, as they are individuated by their subjective, conceptual roles, or equivalently (as far as Loar is concerned), as they are individuated by commonsense psychological explanation. The reason for this, as stated before, is that the two beliefs will interact differently with other beliefs

and attitudes that Paul may have. For example, Paul may come to believe that he has arthritis in his ankles, but not arthritis, and this will be perfectly consistent with his beliefs, as they are individuated by their conceptual roles in Paul's subjective view of the situation (Loar, 1988a, p.104). In contrast, on an oblique univocal reading, the above situation would come out as Paul believing that he has arthritis in his ankles and not believing that he has arthritis in his ankles, which does not seem to make much sense. And Loar is of the opinion that this distinct individuation of Paul's two beliefs comes from them having a different psychological content.

It is at this point that Loar returns to Burge's "arthritis" example, where he claims that:

...Burge's observation that "believes that Bert has arthritis in his ankles" is true of the doctor and Bert on an oblique univocal reading, which I have agreed is correct, does not imply that their beliefs have the same content as that is individuated in commonsense psychological explanation (Loar, 1988a, p.105).

Therefore, Loar is trying to argue thus. With the "Paul" example he tries to show that Paul's *arthritis* and *arthritis* concepts, and so his arthritis and arthritis beliefs, are distinct, due to the different roles they play in Paul's subjective conception of his situation. This is despite the fact that the arthritis and arthritis beliefs are typed as being the same, on an oblique univocal reading. It should be reiterated that when Loar also claims that the beliefs appear distinct, as individuated by "commonsense psychological explanation", I think he is referring to psychological explanation from the subject's viewpoint, i.e., what beliefs and other attitudes the subject would claim were the reasons for his or her consequent behaviour. For instance, in the case of Paul, he would explain his behaviour by reporting that he had distinct arthritis and

arthritis beliefs. Now, when Loar goes back to Burge's "arthritis" example, and considers the arthritis beliefs of Bert (before he consults the doctor) and the doctor, he is trying to argue that this is an analogous situation to that of the "Paul" example. That is, the *arthritis* concepts and arthritis beliefs of Bert and the doctor, are distinct as individuated by their conceptual roles or their psychological content, even though on an oblique reading the beliefs are typed as the same.<sup>9</sup>

### Loar's Response to Objection 2

This objection was that two beliefs which were judged to be the same, as individuated by conceptual role theory, would turn out to be judged as different, as individuated by commonsense ascription. This objection arose from the comparison of Bert's beliefs in actual and counterfactual scenarios. Burge was of the opinion that in the counterfactual world, where arthritis, or rather tharthritis, was considered to be a rheumatoid ailment of joints and bones in general (or of muscle-tissue); Bert's belief, as ascribed by commonsense, would be that he had tharthritis in his thigh. This was in contrast to the situation in the actual world, where arthritis meant a rheumatoid ailment of joints only; here, before he had been to the doctor's, Bert's belief, as ascribed by commonsense, would be that he had arthritis in his thigh. So, it can be seen that this difference manifested itself in different oblique that-clause ascriptions of these particular beliefs, i.e., one that-clause ascribed "tharthritis" while the other ascribed "arthritis".

The power of Burge's example is that the internal, non-relational, facts about Bert are kept the same in the actual and counterfactual situations, and so, the conceptual roles of his beliefs have also been kept the same, across the actual and counterfactual situations. However, despite this identity of non-relational facts and conceptual roles, commonsense ascription still individuates Bert's beliefs as different in the actual and counterfactual situations. As the only difference in the two situations is a social one (the conventional concept of "arthritis"), then Burge concludes that belief ascriptions are individuated and constituted by social factors. Also, according to Loar, because Burge thinks that the contents of beliefs are contained in their oblique ascriptions, then he also thinks that belief-contents themselves, are affected by the social environment, i.e., the content that plays a part in psychological explanation. Therefore, the upshot of this is that because Bert's actual and counterfactual beliefs have different oblique that-clause ascriptions, then they also have different psychological content, which is the assumption that Loar identified in (B). Loar's response to this is to acknowledge that:

...Burge's premise, that our old belief ascription would not then be true of Bert, is correct; and it is an important discovery that belief ascriptions are thus sensitive to social facts which may not be reflected in believers' own versions of things. But the further thesis, that content as it is individuated in psychological explanation depends on independent social factors, is I think not correct (Loar, 1988a, p.106).

So here, Loar is agreeing with Burge that differences in the social environment will affect the truth conditions of belief ascriptions. But importantly, Loar does not believe that the content of an oblique that-clause ascription is what generally plays a role in commonsense psychological explanation. Indeed, he believes that the content captured by an oblique that-clause is the "social content" of a belief. It is this content

that is affected by social factors, usually by the linguistic expression of the belief ascription being deferential to the norms of the linguistic community in which it is expressed (Loar, 1988a, p.110). However, Loar believes that there is only a “loose fit” between the social content of a belief, as expressed in an oblique that-clause, and the psychological content of the belief that is involved in psychological explanation. So, contra (B), even though the oblique that-clause ascriptions of Bert’s beliefs are different in Burge’s example (i.e., “believes that Bert has arthritis...”, and, “believes that Bert has tharthritis...”), this does not necessarily mean that the psychological content is also different (Loar, 1988a, p.106).

In Loar’s view the sameness of conceptual roles of the two beliefs might well be sufficient for sameness of psychological content of the two beliefs. Another way to put the same point, is to suggest that the *arthritis* concept of actual-Bert and counterfactual-Bert is the same, at least from the subjective viewpoints of the Berts involved. Loar comes to this conclusion by again hypothesising on how the beliefs will interact with other beliefs in commonsense psychological explanation. For example, if in the actual and counterfactual worlds, Bert is given an ointment to rub on his arthritic areas, then it seems plausible to envisage that Bert will rub the ointment on his thigh in both worlds. That is, from the subjective viewpoint of Bert, he conceives that he has arthritis in his thigh, both in the actual and counterfactual worlds, irrespective of the social differences between the worlds.<sup>10</sup> As mentioned previously, this subjective conception of how things are in relation to the subject and the world seems to be an important factor in Loar’s account of psychological content.



## 1.2 Assessments and Criticisms of Loar's Narrow Psychological Content

### 1.2.1 *Fregean Puzzles, Interpersonal Beliefs and Psychological Generalisations*

Loar's response to Objection 1 does seem initially plausible, but is it ultimately a successful answer? I will now examine the situation of Bert and the doctor, and try to assess the case for the identity or otherwise of their corresponding arthritis concepts and beliefs. Firstly, it is fairly clear that the situations of Loar's "Paul" example and Burge's original "arthritis" example are not exactly the same. For instance, in the "Paul" example, it is one subject (i.e., Paul) who has the two beliefs with the same oblique univocal reading, but different conceptual roles; whereas, in Burge's example it is two subjects (i.e., the doctor and Bert) who each have one belief with the same oblique univocal reading, but supposedly different conceptual roles. Does this difference between the examples have any great import for Loar's argument?

Pereboom (1995) is of the opinion that this difference does influence the force of Loar's case, at least if Loar believes that a "Fregean puzzle" can help to show the distinctness of the arthritis concepts of Bert and the doctor (Pereboom, 1995, p.426). To make his point, Pereboom first describes a similar type of example to Loar's "Paul" example given above. To summarise Pereboom's example, we imagine that a Frenchman called Alfred, comes to believe (in French) that he has arthrite in his thigh.

Pereboom then claims that "...we therefore attribute to him the belief that he has arthritis in his thigh, using our concept of arthritis – a concept which we also attribute to the experts" (Pereboom, 1995, p.426). Alfred then visits England and discovers (in English) what the experts' conception of arthritis is, i.e., that it is a disease that only affects joints. He therefore comes to believe (in English) that he does not have arthritis in his thigh, but continues to believe that he has arthrite in his thigh (Pereboom, 1995, p.426). However, one day Alfred does make the informative discovery that arthrite = arthritis (which Pereboom calls Frege's test), and this discovery changes many of his inferences, e.g., he comes to believe that he does not have arthrite in his thigh. Pereboom claims that "This is intended to show that the cognitive value of the expert's concept is distinct from that of Alfred's original 'arthritis' concept" (Pereboom, 1995, p.426).<sup>11</sup> However, Pereboom quickly pours cold water over this conclusion, when he declares that:

This type of Fregean puzzle does not...help to show that the cognitive value of Alfred's and the expert's concepts are distinct. One should note that it departs from the form of Frege's original puzzle, since it aims to demonstrate an interpersonal rather than an intrapersonal difference in cognitive value – a difference in cognitive value between Alfred and the expert rather than within Alfred. But such extension of Frege's test would draw distinctions in cognitive value where they do not exist (Pereboom, 1995, p.426).

To prove this last point Pereboom then gives a counterexample. However, before we consider that, let us first ponder what has been claimed in the above. At first sight, it might appear that a Fregean test cannot be involved in Loar's argument for the distinctness of Bert and the doctor's arthritis concepts. Pereboom describes Frege's test for concept distinctness as the following: if a subject has two concepts, A and B, and the identity statement  $A = B$  would be informative for the subject, then the concepts A and B are distinct for the subject. However, in Burge's example, the

situation is that Bert and the doctor both seem aware that they have *arthritis* concepts, e.g., they would both assent to the proposition that Bert has arthritis in his ankles. So this is not a situation where the identity statement “arthritis = arthritis” would be informative to Bert. However, it is worth noting that if Bert was told by the doctor that arthritis is a disease of the joints only, he would come to believe that he did not have arthritis in his thigh. This change in the inferences (and possible behaviour) of Bert, which is related to his concept of, and beliefs about, arthritis, appears to mirror the ones that occur in Loar’s arthrite/arthritis example and Pereboom’s Alfred example.

Moreover, this also shows that there is an identity statement that would be informative for Bert, namely, that “Bert’s conception of arthritis *is not equivalent to* the doctor’s conception of arthritis”; or in Pereboom’s terminology, that “A *is not equivalent to* B” would be informative for the subject (where A and B are Bert’s and the doctor’s *arthritis* concepts). Therefore, in relation to Burge’s original arthritis example, there does seem to be a Fregean test involved in Loar’s argument for the distinctness of Bert’s and the doctor’s *arthritis* concepts.

Indeed, it seems to me that there is the possibility that Loar does not have to describe the situation of Bert and the doctor in the manner of a Fregean puzzle, in order to argue that their *arthritis* concepts could be distinct. He could still argue that despite the two *arthritis* concepts having the same labels and the same oblique *that*-clause ascriptions, the psychological content of Bert’s and the doctor’s concepts/beliefs could still be viewed as being distinct, due to the different inferences and behaviour of Bert in relation to his arthritis concept/beliefs. For example, given several

hypodermic syringes containing cortisone, the doctor would inject Bert in his ankles; however, given the same hypodermics, Bert would inject himself in his ankles *and his thigh*.

Nevertheless, it is now time to examine Pereboom's counterexample to the possibility of using Frege's test in making plausible interpersonal distinctions of cognitive value between two subjects' concepts. For this counterexample, we must first imagine that Alfred is now English and has normal beliefs about the disease he calls "arthritis" in English. Pereboom then declares that:

Suppose Alfred's French counterpart has precisely the same beliefs about arthritis that Alfred has, except that the counterpart would express them in French, and that the inference and behaviour associated with these 'arthritis' concepts are as similar as they can be for people who speak different languages. We would want to make psychological generalizations that regard these concepts as having the same cognitive value. But we could imagine Alfred travelling to France, learning his counterpart's '*arthrite*' beliefs, believing, for a time, that he has two different ailments in his thigh, and later coming to the realization that arthritis = *arthrite*. Hence, this interpersonal Fregean test would illegitimately distinguish the concepts of Alfred and his counterpart (Pereboom, 1995, p.426).

I must admit that this is an interesting and difficult counterexample to fathom, as there are a number of different issues all mixed into it. On the face of it, it seems convincing, but the conclusion to be drawn from it may be more complex than Pereboom supposes. He wants this counterexample to show how Frege's test, when it is extended to making interpersonal distinctions of cognitive value, will make distinctions that are illegitimate, i.e., it will make distinctions of cognitive value between concepts where there are none. However, it is worth noting that in the above example, there does seem to be an *intrapersonal* distinction of cognitive value, between the *arthritis* and *arthrite* concepts within Alfred. Therefore, despite the

*arthritis* and *arthrite* concepts having all the inferential and behavioural similarities that Pereboom mentions (the only difference being a linguistic one), presumably he would judge that it would be a legitimate intrapersonal distinction of cognitive value between the concepts, that would result from Alfred making the informative discovery that “*arthritis* = *arthrite*”.

If this is accepted, then it seems that the illegitimacy of the interpersonal distinction that is made depends on Pereboom’s claim that one could make a psychological generalisation of Alfred and his counterpart’s concepts having the same cognitive value. However, is this a feasible contention? From the subjective point of view of Alfred, the intrapersonal linguistic difference between *arthritis* and *arthrite* seems to be decisive in the resulting difference of cognitive value between the *arthritis* and *arthrite* concepts. However, from the objective point of view, the interpersonal linguistic difference between Alfred and his counterpart is viewed by Pereboom as being irrelevant to making a generalisation that the *arthritis* and *arthrite* concepts have the same cognitive value. On this particular issue, it seems to me that it is wrong for Pereboom to claim that the psychological generalisation he intends involves the cognitive value of the *arthritis* and *arthrite* concepts as being the same. There seems to be a strong intuition that Alfred’s *arthritis* concept and his counterpart’s *arthrite* concept, have distinct cognitive values, due entirely to Alfred calling his concept “*arthritis*” and his counterpart calling his concept “*arthrite*”, and the both of them not realising that “*arthritis* = *arthrite*”.

Indeed, in thinking about this, it seems to me that this situation would also result in the inferences and behaviour of Alfred and his counterpart as not being that similar,

which could be illustrated by using a variation of the earlier mentioned Fregean method for checking whether concepts were distinct. We could ask whether Alfred's and his counterpart's inferences and behaviour would remain the same if their respective *arthritis* and *arthrite* concepts are exchanged between them, one for the other. If the inferences and behaviour of the subjects were different, then we could plausibly claim that the concepts are distinct, i.e., the cognitive values of the *arthritis* and *arthrite* concepts would not be the same.

Using this method, it seems plausible to imagine situations where the resulting inferences and behaviour of Alfred and his counterpart would be different, if their respective concepts were swapped with each other, e.g., if Alfred thought he had arthritis in his shoulder, he would inject some cortisone into it; whereas, if he thought that he had arthrite in his shoulder, he would not inject cortisone into it, and vice versa for his counterpart. Therefore, contra Pereboom, I would claim that his counterexample does not provide plausible evidence that Frege's test would result in an illegitimate interpersonal distinction of cognitive value being made between the respective *arthritis* and *arthrite* concepts of Alfred and his counterpart. For that reason, it also remains feasible that a Fregean test could be used to judge whether in Burge's example, the *arthritis* concepts of Bert (before he consults the doctor) and the doctor, are the same.

Having said this, I do not mean to claim that a psychological generalisation could not be made about Alfred's arthritis concept/beliefs and his counterpart's arthrite concept/beliefs. Considering the inferential and behavioural similarities from an objective viewpoint, then a de dicto univocal reading of the two concepts could result

in a feasible psychological generalisation. What I am claiming is that just because a generalisation could be made, it does not imply, from a subjective viewpoint, that the *arthritis* and *arthrite* concepts have the same cognitive value. Indeed, it seems quite feasible to me that Loar might be able to deal with the general criticism that the narrow contents of beliefs are too fine-grained to provide psychological generalisations between those beliefs, by using the distinction he makes between the social and psychological content of beliefs. The social content of the belief could be used for making wide psychological generalisations, whilst the psychological content of the belief could be used for making narrow psychological distinctions that would explain behaviour from the subject's first-person perspective.<sup>12</sup>

### 1.2.2 *Assessment of Loar's Use of Pierre-Type Examples*

Loar used the Pierre-type examples to provide support for the view that on some occasions, oblique belief ascriptions which are specified by that-clauses, are not fine-grained enough to capture the individuation conditions of the psychological content of a subject's beliefs (Edwards, 1994, p.14; Pereboom, 1995, p.401).<sup>13</sup> This happens when a subject has two beliefs which have the same oblique that-clause descriptions (or are of the same type), but which seem to have distinct psychological contents, as evidenced by the subject's distinct inferences and behaviour when considering one belief and then the other. Indeed, Loar's point that the London-belief and the Londres-belief appear distinct in relation to psychological explanations of inference and behaviour, also ties in with the original point Kripke was making about the Millian supposition that "...coreferential proper names make the same contribution to

content” (Pereboom, 1995, p.404).<sup>14</sup> If, as Kripke claims, we follow the standard rules of belief ascription (involving translation and disquotation), then we will end up with Pierre having contradictory oblique, univocal, beliefs ascribed to him (i.e., “London is pretty” and “London is not pretty”), despite the fact that we can consider Pierre to be rational in his thinking. This problematic result arises only if we assume, along with Mill, that coreferential proper names do make the same contribution to the content of a belief (Pereboom, 1995, p.404).

#### What Conclusions Can be Drawn From Loar’s Use of the Pierre-Type Examples?

Nevertheless, one strategy of philosophers critical of the above views (who are generally of an externalist persuasion) has been to question what exactly Loar’s “Pierre” and “Paul” thought experiments are supposed to show about the relationship between that-clauses and psychological content, and whether what is shown is strong enough to support Loar’s position. For example, this is the approach adopted by Biro (1992), who is unsure exactly how strong a position Loar wants to defend, e.g., does Loar want to argue that that-clauses can *sometimes* capture the psychological content of a subject’s belief, or does he want to say that that-clauses can *never* capture the psychological content of a belief? How strong a position does the phenomenon of Kripke’s “Pierre” example actually support? (Biro, 1992, p.284) After some deliberation, Biro comes to the conclusion that Loar wants to support a “radical” position, where that-clauses in general never capture the psychological contents of beliefs.<sup>15</sup>



The question that Biro then asks is what support does this radical position get from Kripke's "Pierre" example. After examining several theses, Biro comes to the conclusion that the Pierre example would only be able to support something as strong as the following claim, "For no belief can the obvious that-clause be guaranteed to capture its psychological content": which basically means that he believes the Pierre example does not provide support for Loar's radical thesis that that-clauses *never* capture the psychological content of beliefs (Biro, 1992, pp.287-288). Biro's thinking behind this conclusion is that even though the "obvious that-clause" (i.e., a straightforward disquotational one) might not capture the psychological content of a complicated belief like Pierre's, it seems quite plausible that circumlocution using further that-clauses may well be able to capture the psychological content of Pierre's belief (Biro, 1992, p.287). Furthermore, he also criticises Kripke's use of the general principles of "disquotation" and "translation" in ascribing beliefs to Pierre in such a complicated case.<sup>16</sup> For instance, I suppose that Biro means that we could quite coherently ascribe the beliefs that "Londres is pretty" and "London is not pretty" to Pierre, while adding the caveat that Pierre does not realise that Londres = London, and so he consequently may not make the correct inferences with regard to the Londres/London beliefs that he has.

I think that Biro is correct when he claims that Kripke's "Pierre" example does not give support for Loar's radical position, i.e., the view that that-clauses would never capture the psychological content of a belief. It seems plausible to me that through circumlocution using further that-clauses, the psychological content of Pierre's Londres/London beliefs could eventually be contained in those aforementioned that-clauses. Indeed, Loar identifies two positions that one can take on the issue of

psychological content and that-clauses; a radical one (which Biro mentions) and a moderate one (Loar, 1988b, p.121). As stated previously, the radical position would be that oblique that-clauses never capture the psychological content which individuates beliefs in commonsense psychological explanation. In contrast, the moderate position is that on some occasions, that-clauses might well be able to capture the psychological content which individuates a belief.

For example, in the actual world, if Bert believed that he had arthritis in his ankles, and also had the correct conception of arthritis, i.e., that it is a rheumatoid ailment of the joints; then I think it is fair to say that when the belief “that Bert has arthritis in his ankles” is ascribed to him in an oblique that-clause, then the content of this that-clause would be the same as the psychological content of Bert’s belief. Furthermore, if a doctor examines Bert, and also comes to believe that he has arthritis in his ankles, then in this case, the same oblique readings of Bert’s and the doctor’s beliefs, would mean that their beliefs also shared the same psychological content. It seems plausible that when the objective, social, conception of an arthritis-belief, as contained in a that-clause, is the same as the subjective, individual, conceptual role of the arthritis-belief, then sameness of that-clause ascription in two beliefs might well signify sameness of psychological content of the two beliefs.

However, despite the above, it is important to understand when and why Loar uses Kripke’s “Pierre” example and his own variation of it. In his 1988a, Loar does claim that he will be arguing:

...that psychological content is not in general identical with what is captured in that-clauses, that commonsense constraints on individuation induce only a loose fit between contents and that-clauses...(Loar, 1988a, p.102).

Now, this could well be viewed as a fairly strong thesis, but it is certainly not as strong as Loar's radical thesis that that-clauses never capture the psychological content of a belief. Indeed, in using the "Pierre" example (and also his "Paul" example), Loar is principally attacking the view, as put forward in (A), that sameness of oblique belief ascription is sufficient for sameness of psychological content. In saying this, Loar might not mean that an oblique ascription would *never* capture the psychological content of a belief. In rejecting (A), he is only rejecting the view that there is an *implication* that the sameness of oblique belief ascription will lead to sameness of psychological content of the beliefs. With his use of the "Pierre" variation, I think that Loar is successful in his aim of casting doubt on the plausibility of (A). In the variation, Pierre has the belief that "Londres is pretty" and the belief that "London is pretty". On an oblique univocal reading, these two beliefs are the same, but it seems fairly clear that the psychological content of the two beliefs are not the same, i.e., as the beliefs would be individuated in the psychological explanation of the inferences and behaviour of the subject concerned.

Of course, Biro's point could be made here: that is, with further circumlocution, the psychological content of Pierre's Londres/London beliefs could eventually be captured by that-clauses. However, I think that this would be missing the point that Loar is trying to make here. What he is claiming is not just that the psychological content of Pierre's Londres and London beliefs are distinct, Loar is also claiming that they are tokens of distinct belief-types, which would basically mean that sameness of oblique belief ascription is not *sufficient* for sameness of belief (as the latter is

individuated by narrow content). This, in turn, would mean that Pierre's beliefs are not just being individuated by their oblique that-clauses, but also by the subjective conceptions of Pierre, that is, by how Pierre is conceiving the situation or state of affairs in the world to be. In this particular case, Pierre is of the opinion that there are two distinct cities in the world, called Londres and London, both of which are pretty.

It should also be kept in mind that Loar is making this point in particular with Burge's "arthritis" thought experiment in mind, especially Burge's first objection to the notion of conceptual role individuation of beliefs (see Loar's response to Objection 1, where he uses the arthritis/arthritis example to reinforce the point he is making). Indeed, it seems that the anti-individualist conclusions that Burge draws from his arthritis example do depend on singular oblique or de dicto that-clauses being viewed as having a certain content in actual and counterfactual situations. If Loar can cast some doubt on what the content of these that-clauses are, then he could in turn cast some doubt on the anti-individualist conclusions that Burge comes to.<sup>17</sup>

### A Difference of Concepts, But a Type-Identity of Beliefs

Another strategy to dilute the force of the internalist intuitions which come from the "Pierre" example, is to accept that the conceptions of Pierre (both in Kripke's original version and Loar's variation of it) are distinct, but to question what exactly this result amounts to in relation to the beliefs that Pierre has. As we have seen already, the notion that Pierre has distinct conceptions does seem fairly compelling, for a number of related Fregean reasons. Firstly, there is what Evans (1982) called "The Intuitive

Criterion of Difference”, which is a Fregean method for discerning the identity of “senses”: “If a subject will assent to an expression M but not to an expression N then M and N have distinct senses” (Evans, 1982, p.18; Edwards, 1994, p.162).<sup>18</sup> This identity criterion of conceptions can be adapted to explain the differences in a subject’s behaviour or actions, as Edwards (1994) states:

...where M and N are expressions employed to represent an agent, S’s, conception, at time t, of a state of affairs A, if S acts differently when he entertains M than when he entertains N, then M and N represent distinct conceptions of A. Put differently, it is not the case that M and N are identical characterisations of S’s cognitive state at t, if they result in distinct actions when the states which M and N characterise are exchanged one for the other (Edwards, 1994, p.162).

I suppose that it could be claimed here that M and N would have to be in languages the subject understands, so that the above criterion would not apply to Pierre-type examples. I’m not sure about this, as Loar does claim at one point that “...”London” is univocal as we use it” (Loar, 1988a, p.102), and also that in relation to Pierre’s apparently distinct London and Londres conceptions:

The point does not depend on translation; parallel cases arise in which someone mistakenly thinks a name names two things, and ascribes the same predicate twice (Loar, 1988a, p.103).

So I suppose that the point here is that even though different languages are involved, the principle of the Intuitive Criterion of Difference is still the same, just as if the example only used one language which the subject fully understood. Therefore, for example, in Kripke’s original Pierre thought experiment, if Pierre would assent to the expression “Londres is pretty” but would not assent to the expression “London is pretty”, then by Frege’s intuitive criterion of difference, we should judge that, for

Pierre, the expressions have distinct senses or conceptions. Furthermore, if the behaviour of Pierre appears to be different for his Londres and London beliefs (e.g., he rushes out and buys some cheap air tickets to Londres, but does not do so if the cheap air tickets are to London), then we can again conclude that his conceptions of Londres and London are distinct. Therefore, there are powerful Fregean considerations with regards to the identity of a subject's concepts which seem to point towards Pierre's concepts of Londres and London as being distinct.

In examining the above claim I will look at Edwards'(1994) treatment of Loar's "Pierre" variation. Edwards does not use the Londres/London example, but uses his own version which involves natural kind beliefs about "tigres" and "tigers". It is supposed that while in France, S comes to have beliefs about animals called "tigres" (these are only descriptive beliefs, as S has had no perceptual experience of a tigre), which he comes to believe "...are wild, run free, have black and orange stripes, but are never found in zoos" (Edwards, 1994, p.213). It is further supposed that S will assent to the expression "Tigres are never found in zoos". What happens next is that S moves to London and gets a job at London Zoo, where he becomes perceptually acquainted with animals called "tigers", although he does not realise that these tigers are the same animals which he calls tigre from his time in France (Edwards, 1994, p.213).

Therefore, at this point, what can we say about S and his tigre/tiger conceptions? Well, as Edwards suggests, it seems feasible that S would assent to the expression "Tigers are found in zoos" but dissent from the expression "Tigres are found in zoos", despite the fact that both expressions express the same proposition; or as Loar would

put it, despite the fact that both belief ascriptions have the same univocal de dicto occurrence of a general term in their that-clauses. Therefore, according to the intuitive criterion of difference, it would seem that S's senses or concepts of *tigre* and *tiger* are distinct. Also, according to Frege's test, it seems plausible that "tigre = tiger" would be informative for S, so S's *tigre* concept and *tiger* concept could again be considered to be distinct. Edwards also brings in inferential and behavioural elements in relation to the identity of S's concepts, with S having promised his mother a photograph of a tigre, but while at London Zoo he does not take a photograph, despite coming into contact with tigers every day (Edwards, 1994, p.213). This inferential failure and corresponding lack of certain behaviour of S, again leads one to conclude that S's concepts of *tigre* and *tiger* are different.

Having introduced the example, Edwards states that:

...[t]he challenge presented to type externalists by such a case is to say why factors beyond the body of the subject have an explanatory role to play in accounting for the behaviour of the subject (Edwards, 1994, p.213).<sup>19</sup>

At another point, he is more specific about this challenge, where he claims:

In this kind of case the externalist has to show why the thoughts of the subject whilst in London and in France should be regarded as being of the same type in spite of differences in the E-components of the thoughts (Edwards, 1994, p.214).<sup>20</sup>

Therefore, the challenge for the externalist is to show that even though S appears to have distinct beliefs about tiges and tigers, they should still be regarded as being of the same belief-type, i.e., as tiger beliefs (or in the case of Pierre and his Londres/London beliefs, they should be typed as London beliefs). As can be garnered

from the above, Edwards' strategy for accomplishing this end is going to be the following: he is going to argue that S's tigre and tiger beliefs have distinct E-components, but that despite this, the beliefs should still be typed by the same natural kind "tiger" which they all concern. It can be seen then that Edwards can concede that in Pierre-type examples, a subject may have distinct conceptions of the same object or natural kind; indeed, he says that Loar has a "solid motivation" for his claim that the relevant concepts of a subject in such a thought experiment are distinct (Edwards, 1994, p.216). However, before he concedes this point, Edwards does briefly question whether Loar is right to claim that Pierre's beliefs that "Londres is pretty" and "London is pretty" "...are as distinct as my beliefs that Paris is pretty and that Rio is pretty" (Loar, 1988a, p.103). He believes that if Loar's reasoning is extended to natural kind thoughts, then:

...Loar's position would be one in which S's beliefs about what he calls "Tigres" and what he calls "Tigers" could be said to be as distinct as his beliefs about tigers and mercury...such a claim seems highly implausible. First, it does not appear possible to construct a situation which is appropriately analogous to that under discussion here in which the kind of mistake which Loar conceives to be possible is in fact feasible. The reason why is that as these types of examples are constructed it is required that a subject assent to and dissent from two propositions which most competent bilinguals would consider adequately close translates of each other... (Edwards, 1994, p.215).

Now, I think that some of the above points are debatable. Firstly, I am not sure whether Loar's claim that Pierre's Londres/London beliefs are as distinct as Paris and Rio beliefs, is equivalent to claiming that S's tigre/tiger beliefs are as distinct as beliefs about tigers and mercury. It will be noted that Paris and Rio could be said to be entities of the same type, i.e., cities, but that they are different. Therefore, what I think Loar is trying to say is that according to Pierre's conception of the world, there are two distinct cities, Londres and London, and that for Pierre, they are as distinct as



beliefs about the cities Paris and Rio. I don't think this is too implausible, and is not analogous to claiming that S's beliefs are as distinct as beliefs about tigers and mercury. The latter comparison seems to be dealing with beliefs about different types of natural kinds, i.e., tigers and mercury, whereas, Loar's comparison is basically dealing with objects of the same type, i.e., cities. Perhaps if Loar had claimed that Pierre's Londres/London beliefs were as distinct as beliefs that "Paris is pretty" and "The Moon is pretty", then possibly it could be said that this was analogous to the case of tiger and mercury beliefs. However, based on Loar's original comparison, it seems more plausible to claim that S's tigre/tiger beliefs are as distinct as, say, lion and tiger beliefs, or leopard and cheetah beliefs, etc.

The other point that Edwards makes about examples of this kind, is that the terms involved are close translates of each other, and would be spotted as such by competent bilinguals; with the result that it would not in fact be possible to really construct an example of this kind. Now, I agree that if the example hinged on competent bilinguals assenting to, and dissenting from, expressions containing the terms "tigre" and "tiger", then this would not really be feasible, as the bilinguals would be able to notice that the key terms in the expressions were close translates of each other, and so were expressing the same proposition.

However, I think that some of the details of this example can be changed to make it more feasible. For example, in the Pierre thought experiment, Pierre is not actually a competent bilingual, but is actually a monolingual French speaker who lives in France (Kripke, 1979, p.119; Loar, 1988a, p.102). Accordingly, we could imagine that S was also in this situation, he came to learn about animals called tigres, and would assent to

the expression (in French) of “Tigres are not found in zoos”. He consequently moves to London and starts work at London Zoo, where he becomes acquainted with animals which are referred to as “tigers” in English. He still does not speak much English, but has enough knowledge to be able to assent to the expression “Tigers are found in zoos”.

Now, this scenario seems fairly plausible to me. If it was claimed in response to this that the assenting or dissenting involved had to be “knowledgeable” or “reflective” in nature, then there are other options one could take. For instance, there are other strategies for discerning the identities of concepts that a subject is using. Assenting to, and dissenting from, certain expressions, is the intuitive criterion of difference method for discerning whether a subject has distinct conceptions. But one could apply this method to the behaviour of the subject in question (as Edwards stated above), so there would be no assenting or dissenting to expressions involved in judging whether a subject has different conceptions of an object or state of affairs: one would be able to decide this point by observing differences in the behaviour of the subject, depending on whether he/she is entertaining, say, tiger beliefs or tigre beliefs at a time  $t$ , or put another way, the subject’s behaviour would not remain constant if his/her tigre beliefs were exchanged for tiger beliefs. Moreover, there is also Frege’s test for checking for differences of concept, e.g., it is plausible that the identity statement “Tigre = Tiger” would be informative for  $S$ , so that the two terms would be judged as involving different concepts, at least as far as  $S$  was concerned.

Of course, it is important to note that Loar could have made his point that sameness of oblique belief ascription does not imply sameness of psychological belief content, by

not involving terms which are close translates of each other. For example, as quoted earlier, he could have constructed an example where a subject mistakenly believes that a general term or a name names two things, with the result that the subject could be ascribed two beliefs with the same de dicto predicates. With a general term, there could be a situation where a subject correctly believes that “gazebo” is a term for a rather posh summerhouse, but also mistakenly believes that “gazebo” is a term for a baby gazelle. With this situation, there could be the same oblique gazebo belief ascriptions, which had radically different psychological contents for the subject concerned. With a name, there could be something similar to Kripke’s “Paderewski” example, where a subject may not realise that Paderewski the statesman is the same person as Paderewski the pianist, and so may very well be ascribed the same oblique Paderewski beliefs, but, despite this, as far as the subject is concerned, each belief has a different psychological content.<sup>21</sup>

Nevertheless, as stated above, Edwards does acknowledge that there is a “solid motivation” for the view that S’s *tigre* concept and *tiger* concept are distinct. Of course, what he wants to argue is that this difference of concepts does not amount to S having distinct belief-types (as Loar claims). Edwards can claim that the distinctness of S’s *tigre* and *tiger* concepts can be accounted for by the distinctness of the E-components of S’s *tigre* and *tiger* beliefs (Edwards, 1994, p.216). This enables him to claim that:

...it can be argued it is the fact that the intentional states of S concern the same natural kind that is most relevant to the typing of those states. This warrants the claim that S’s conceptions of Tigers only serve to explain his failure to photograph one when it is accepted that his conceptions are of one kind of object. This indicates that explanation of S’s failure to act involves tacit reference to tigers...If internalism is true no reference to tigers is required in

explanation of the explanandum event (S's not taking the photo); but if (type) externalism is true referential components of S's psychological states do perform explanatory work (Edwards, 1994, p.216).

Before examining Edwards' reasoning above, I should say that I have the distinct feeling that the lessons of Pierre-type thought experiments will probably not provide knockdown arguments against either internalist or externalist positions on intentional content. What seems to be involved here are intuitions about the correct principles of psychological explanation of human behaviour. The Pierre-type examples are usually understood as providing strong intuitions for the internalist position on psychological explanation, but as we can see, Edwards is trying to counter these internalist intuitions by providing some externalist intuitions about psychological explanation.

The present point at issue, over which all these intuitions are swirling, is how to explain S's behaviour at London Zoo, i.e., to explain why S does not take a photograph of a tiger. It can be seen that Edwards' explanation for S's inaction does seem quite plausible, while it also keeps the belief-types of S the same, i.e., both beliefs are considered to be tiger-beliefs (so making the explanation conducive to externalism). However, this explanation does raise a puzzling issue, that is, what exactly are the roles of the distinct E-components of S's beliefs in this explanation, what purpose are they supposed to serve, if any? Perhaps Edwards believes that they play no role in the psychological explanation of S's behaviour, but this view does not seem very compelling. Maybe he intends the distinct E-components to play some sort of individuating role, but that the most relevant individuating role in explaining S's behaviour is carried out by the typing of his two beliefs as being concerned with the same natural kind, i.e. tigers.<sup>22</sup>

Be this as it may, it seems to me that there is a tension in Edwards' externalist psychological explanation of why S does not take a photo of a tiger at London Zoo. I think that the tension can be best illustrated in terms of the Nagelian idea of subjective and objective points of view. The issue seems to be whether, when we are ascribing beliefs to S, in order to explain his/her behaviour, we adopt the subjective viewpoint of S him/her-self, or take an objective viewpoint of his/her situation. By this I mean the following. If we adopt the subjective viewpoint of S, this means that we effectively take up S's point of view on the world, and try to work out what his/her actual beliefs or other propositional attitudes are from the first-person perspective, in order to explain his/her behaviour. Now, it seems to me, that if we follow this strategy, there is going to be a strong intuition that S's behaviour is going to be explained by him/her having two distinct types of beliefs, i.e., tigre beliefs and tiger beliefs. S promised mother a photo of a tigre, but when at London Zoo, S believes that the animals that are called tigers there are not the same animals as tigres, and so he/she does not take a photo of one. So, from S's subjective viewpoint, the most relevant typing of beliefs to explain his/her behaviour is going to involve distinct tigre and tiger beliefs.

In contrast to this, if we adopt an objective point of view on S's situation, this means abstracting away from S's subjective viewpoint when giving an explanation of his/her behaviour, i.e., it is as if we are taking in the whole picture of what is going on, including factors that S does not know or realise. Now, when we take this viewpoint, then it does seem plausible to claim, as Edwards does above, that the relevant factor in explaining S's behaviour is that he/she does not realise that "tigre = tiger",<sup>23</sup> i.e., that both his beliefs involve tacit reference to the same natural kind Tiger. However,

while this might seem plausible as an explanation of S's behaviour from an objective, third-person point of view, it seems totally implausible in explaining S's behaviour from the subjective viewpoint of S him/her-self. From S's viewpoint, it is his/her belief that tigras are different creatures from tigers, that is the most relevant factor in explaining why he/she does not take a photo of a tiger,<sup>24</sup> i.e., contra Edwards, no reference to an "objectual component" such as the natural kind Tiger, is needed to explain S's behaviour.

A "privileging" of the objective stance also seems to be present when Edwards later considers "...the conditions under which content attributions occur" (Edwards, 1994, p.217); this involves a "radical interpreter" who ascribes to subjects "...thoughts about objects with which they commonly interact" (Edwards, 1994, p.217). This seems a very plausible strategy, however, it is interesting to note that the interpreter seems to be presupposing an externalist position, by ascribing to a subject thoughts about objects that he/she *should* be having, given his/her current environment. The notion of a radical interpreter seems to essentially involve adopting an objective point of view on a subject's situation, when making content attributions to that subject.

By saying this, I mean that this strategy does not seem to take into account how the subject's conceptions of his/her current environment may be affecting the individuation of his/her thoughts, at least with regard to how the subject's thoughts interact with each other in a psychological explanation of his/her behaviour. Edwards does admit that with the situation of S at London Zoo, it is not going to be unproblematic for the interpreter to ascribe thoughts to S that explain his/her behaviour (i.e., his/her not taking a photo of a tiger). However, it is worth pointing

out that this is only going to be problematic for the interpreter, who appears to be of an externalist persuasion in regards to making content attributions. Edwards suggests that the interpreter gets to know S and his/her history a little better, so that he can come to appreciate that S does not realise that “tigre = tiger”, and that S is really referring to the same animal when he/she talks of tiges and tigers (Edwards, 1994, pp.217-218), which as far as Edwards is concerned “...is the source of the puzzle” (Edwards, 1994, p.218).

However, it can be seen that once again the interpreter is adopting an objective third-person viewpoint on ascribing thoughts to S that explain his/her behaviour, as he is taking into account a factor that S does not know from his/her first-person subjective viewpoint, and so will not have a thought about. In order to explain S’s behaviour, it seems far more intuitive to me to try to adopt the subjective viewpoint of S, to ascribe thoughts or beliefs to him/her, that, from his/her viewpoint, actually are affecting his/her behaviour, and not just ascribe thoughts that *should* be affecting his/her behaviour (according to the externalist).<sup>25</sup> To achieve this, perhaps the disquotation principle of ascribing beliefs could be used, i.e., have S assent to, and dissent from, various tigre expressions and tiger expressions. If this strategy is adopted, it seems plausible that the outcome would be that the reason S did not take a photo of a tiger would be that he/she believed that the animals at London Zoo were not tiges.

Consequently, it seems to me that the tension in Edwards’ above explanation of S’s behaviour, that adopting subjective and objective viewpoints brought out, is the following: it seems strongly unintuitive and implausible to allow that S has distinct *tigre* and *tiger* concepts, but then to deny that this factor is the main or relevant one

that contributes towards the typing of S's beliefs in order to explain his/her behaviour.<sup>26</sup> Therefore, even though in most cases, the belief ascriptions from the subjective and objective perspectives will probably be the same, in cases where there is a divergence of ascriptions, it seems more plausible to me to follow the subjective belief ascriptions in explaining the subject's behaviour.

Therefore, in this chapter I have introduced Loar's notion of narrow psychological content and tried to defend it from various criticisms. In the next chapter I will evaluate this notion and examine how Loar intends to provide a specification for it, particularly in relation to Burge's "arthritis" thought experiment.

## Notes to Chapter 1

<sup>1</sup> To illustrate CRS, consider the following example. Imagine that a person is faced with a kangaroo and a giraffe, and is asked to pick-out the kangaroo. This the person duly does by pointing at the kangaroo and uttering "That is a kangaroo". From this we can claim that the person does seem to have the concept of "kangaroo", and so for him, the meaning of the term "kangaroo" is *kangaroo*, as this is the concept that the term signifies. That is, because of the appropriate inferential relations that the person's kangaroo-concept has with his sensory inputs, other concepts/thoughts and the behavioural output of pointing and uttering, means that the person does have the correct concept of "kangaroo".

Many commentators see CRS as the narrow, "in the head", component of a two-factor theory of content; with the wide component of the theory usually being comprised of some sort of referential truth conditional semantics (Block (1986) p.627. For examples of such two-factor theories of content; see Field (1977), McGinn (1982), Loar (1982), Block (1986)). The basic idea behind these two-factor theories is that it takes both the narrow and wide components to give the complete thought-content. This seems plausible, as thoughts do seem to have a narrow content and a wide content. The narrow content of a thought could come from the conceptual role of the representational state involved with the thought, in relation to other thoughts, and with sensory inputs and behavioural outputs. This narrow conceptual role content would seem to depend on how the subject is representing the world, and not on the actual external world connections those representations have (McGinn (1982) p.210; Block (1986) p.627). Moreover it is the conceptual relations between thoughts that have the dominant effect on producing and explaining the consequent behaviour of the subject. However, thought-content is also individuated according to its truth conditions, which depends on what the actual external world referents of the thoughts are. This is where the wide referential truth conditional content of the thought comes into play, which is deemed to be the thought's proper semantical content (McGinn (1982) p.210). Even though the narrow conceptual role component may individuate a thought differently from the wide referential truth conditional component, as McGinn claims "...both are needed to fix [the] total content" (1982, p.211) of the thought.

<sup>2</sup> Where "what it would be like" is used in the same sense as in Nagel (1974).

<sup>3</sup> By "individualist" concepts, I mean those concepts as they are understood from the subject's first-person perspective, how they inferentially interact with the subject's other concepts and beliefs.

<sup>4</sup> Even though Loar concentrates mainly on issues that Burge raises in connection with that-clauses and the individuation of propositional attitudes, he also mentions an example of Stich (1983), which also seems to tie the individuation of a belief to the oblique that-clause which describes it. In Stich's



example, a blind person and a sighted person are each told there is a cat in the next room, from which Stich argues that we should intuitively ascribe the same belief, on a *de dicto* or oblique reading, to both sighted and unsighted persons, i.e., the belief that there is a cat next door (Loar, 1988a, p.100). Stich claims that our intuitions would stay the same, even if the blind person was replaced by Helen Keller, whose perceptual deficiencies were greater than the blind person's. That is, we would ascribe the same belief to Helen Keller and the sighted person, using an oblique *that*-clause, i.e., they both believe that there is a cat in the next room. In this way it can be seen that Stich is tying the individuation of the belief, and so its "content", to what is ascribed in the oblique *that*-clause.

<sup>5</sup> Loar (1987) refers to "representational" content rather than "psychological" content. I think that in Loar's view the two contents are interchangeable, as a subject's psychological content is how the subject conceives of the world, that is, how the subject represents to himself how things are in the world.

<sup>6</sup> It is a similar story with the example of Stich's, where (A) seems to apply to the reasoning there, irrespective of the conceptual roles and cognitive elements of the beliefs involved.

<sup>7</sup> With regards to the term "univocal", I take this to mean that the words contained in both belief ascriptions are understood to have their conventional meanings. So, for example, the word "Londres" couldn't mean something silly like "cabbage" etc.

<sup>8</sup> Loar also uses a variation of this thought experiment to counter Stich's earlier claim that recognitional factors or abilities are not crucial in individuating beliefs.

<sup>9</sup> Segal (2000) also tries to illustrate a clear-cut difference of "arthritis" concepts and beliefs between Bert and the doctor, using a strategy similar to that Loar uses in his arthrite/arthritis example (2000, p.n158). The method that Segal uses to circumvent the problem of Bert and the doctor both appearing to have concepts of *arthritis*, is to bring in a synonymous word or expression for the concept of arthritis. In this case, it is considered that the expression "inflammation of the joints" is synonymous with the meaning or concept of the term "arthritis". Segal's reasoning then goes as follows. In the case of the doctor or expert, even though there are two expressions, i.e., "inflammation of the joints" and "arthritis", because the expert will know that these expressions are synonymous, they will both express only one concept as far as the expert is concerned (2000, pp.65-66). However, it seems logically plausible to suggest that this will not be the case for Bert (Segal calls his character Alf). Even if it is granted that, through partial understanding and deference to the social/linguistic environment he is in, Bert can be ascribed arthritis beliefs, the concept that his term "arthritis" expresses cannot be the same as the concept that the expert expresses with his use of the expression "arthritis".

The reason for this claim is that even though Bert has the belief that he has arthritis in his thigh, he would presumably not have the belief that he had an inflammation of the joints in his thigh, as he knew that his thigh was not a joint; therefore, it seems plausible to claim that for Bert, the expressions "arthritis" and "inflammation of the joints" are considered to express different concepts (2000, p.65). Therefore, as Segal sees the situation, this is a case, where, given the expressions "arthritis" and "inflammation of the joints", the expert believes that they express one and the same concept, whereas Bert believes that they express two different concepts; with it being more probable that it is the term "arthritis" which is expressing different concepts, as used by Bert and the expert (2000, p.66). The similarity of Segal's example and Loar's arthrite/arthritis example seems fairly clear, as Bert could doubt that he had an inflammation of the joints in his thigh and still believe that he had arthritis in his thigh (while still remaining consistent in his reasoning), just as Paul could doubt that he had arthritis in his thigh, but still believe that he had arthrite in his thigh.

<sup>10</sup> Cummins (1991) comes to roughly the same conclusion as Loar, where he claims that even though the semantic content of two beliefs are considered different, this does not entail that their psychological contents are also different (1991, p.63). For Cummins, the distinction between the psychological content and semantic content of a concept is that "...[t]he psychological content is the knowledge in (or accessible via) the relevant data structure: the semantic content...is the sort of thing that enters into truth conditions. I'll call this view of concepts the encyclopedia view of concepts to emphasize the idea that the psychological content of a concept is like an encyclopedia entry rather than a dictionary entry (i.e., an entry that specifies a meaning)" (1991, p.63). Perhaps one could say here that Cummins' encyclopedia view of concepts is similar to Loar's conceptual role view of psychological content. That is, with both views, the ultimate psychological meaning of a concept is the way that it interacts with the subject's other concepts or "encyclopedia entries", which is distinct from the semantic or social meaning that the concept has.

<sup>11</sup> I take Pereboom's "cognitive value" to be synonymous with Loar's "psychological content"; it is the content-like aspect that a thought has as it interacts with other thoughts, and with sensory inputs and behavioural outputs.

<sup>12</sup> For example, this distinction between social and psychological content could deal with a similar example to Pereboom's, this time given by Rechenauer (1997), which he thinks illustrates a problem for Loar's views on the individuation of psychological states and their respective contents, which involves the plausible constraint of intersubjective shareability of psychological states (1997, p.59). In this case, the intersubjective shareability of psychological states could be captured using the social content of those states, whilst still allowing the psychological content to make more fine-grained distinctions between those states.

<sup>13</sup> Conversely, an externalist criticism of narrow content theories is that the narrow content postulated is not coarse-grained enough to provide the generalisations that are required for psychological explanations.

<sup>14</sup> Or, as Kripke (1979) puts it "...it would appear that proper names of the same thing are everywhere interchangeable not only *salva veritate* but even *salva significatione*: the proposition expressed by a sentence should remain the same no matter what name of the object it uses" (1979, p.104).

<sup>15</sup> As Biro notes, Loar (1988b) does declare at one point that he wants to give the radical position "a run for its money" (1988b, p.121).

<sup>16</sup> In relation to belief ascription, Kripke defines the Disquotational Principle as that "If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that p" (1979, p.112-3); Biro expresses the Disquotational Principle as "...according to which sincere assent to a sentence warrants ascription of a belief with the content that sentence expresses" (1992, p.282). The Translation Principle, according to Kripke, is that "If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language)" (1979, p.114). In ascribing a belief using the principle of disquotation, the consequences are of interest, especially if it is applied to the question of what belief we should ascribe to counterfactual-Bert, in Burge's "arthritis" thought experiment. That is, would Bert reflectively and sincerely assent to a sentence stating that he (Bert) had "tharthritus" in his ankles? Perhaps I will return to this at a future point.

<sup>17</sup> I think this point is similar to the one that Rechenauer (1997) makes, based on a discussion he had with Loar himself. He declares that "...Loar has a simple reply. He took principles (A) and (B) for the premises used by Burge and others in order to justify the inference from the broad content of that-clauses to the broad content of the mental kinds ascribed by them. If the antiindividualist does not need them...then he seems obliged to supply some other principle which guarantees the inference from that-clause content to psychological content" (1997, p.54).

<sup>18</sup> This criterion for determining the identity of a subject's senses is of course related to what Pereboom (1995) called "Frege's test", which states that: "In general, if a subject has the concepts 'A' and 'B', and if 'A = B' would be informative for her, then 'A' and 'B' are different concepts for that subject, and thus for her they differ in cognitive value" (1995, pp.402-403). It can be seen that there is a powerful intuition that for Pierre, the identity "Londres = London" would be informative to him, and so according to Frege's test, the concepts of "Londres" and "London" would be considered different for Pierre.

<sup>19</sup> Edwards defines "type externalism" as the "...thesis which claims that externalism is true for at least some of the types of psychological states recruited in explanations of action" (1994, p.xi).

<sup>20</sup> The "experiential component" or "E-component" is an idea introduced by Edwards in order to supplement the that-clauses which describe demonstrative thoughts: this is because that-clauses are viewed as being inadequate for specifying the content of demonstrative thoughts. The E-component will add experiential and cognitive elements to the individuation of the content of a demonstrative thought. It can be seen that just as Edwards used his E-component to supplement the that-clauses describing demonstrative thoughts, he is now attempting to use his E-component to supplement the that-clauses of belief ascriptions. The E-component of Edwards will make another appearance later on in this thesis, when the discussion is concerned with Loar's notion of realization conditions.

<sup>21</sup> As Pereboom puts it, for the subject, the identity statement "Paderewski = Paderewski" would be informative (1995, p.404).

<sup>22</sup> I don't think that Edwards would want the E-components of S's beliefs to have too much individuating power, as this position could allow Loar's thesis about that-clauses and psychological content to survive. It could be claimed that despite the sameness of oblique that-clauses of S's tiger beliefs, this does not imply that the psychological content of those beliefs are also the same. This claim depends on equating the "content" of S's distinct E-components from each of his tiger beliefs with Loar's notion of psychological content.

<sup>23</sup> This same point could be made slightly differently using the phraseology of Patterson (1990), where the relevant factor would be that S was ignorant of the “metalinguistic truth” that “tigre” and “tiger” have the same meaning (1990, p.320).

<sup>24</sup> That is, as individuated by commonsense psychological explanation, it seems crucial that S has distinct belief types.

<sup>25</sup> By this I mean that the externalist position seems to allow that from the subjective viewpoint of the subject, his/her having distinct tigre and tiger beliefs seem to explain his/her behaviour at the zoo. However, the externalist will claim that the main and most relevant typing of the subject’s beliefs that offer an explanation of his/her action, will be that his/her “distinct” beliefs are actually type-identical, that is, they both concern the same natural kind Tiger, but that the subject does not realise this. I think that this point is similar to the one that Patterson (1990) makes, where she claims that externalists are making a distinction at the level of second-order beliefs, to explain the subject’s behaviour, with the aim of keeping the first-order beliefs of the subject type-identical (1990, p.321).

<sup>26</sup> A similar objectifying strategy for dealing with the internalist intuitions of Kripke’s “Pierre” example is given in Frances (1999). Frances’ example uses a situation where it is mistakenly believed that the name “Hellman” refers to two different people, but this example also has the same subjective and objective tension in it that Edwards’ example has.

## CHAPTER 2

### AN EVALUATION OF LOAR'S PSYCHOLOGICAL CONTENT IN RELATION TO BURGE'S "ARTHRITIS" THOUGHT EXPERIMENT

Having examined Loar's responses to the two objections that Burge raised against conceptual role individuation of beliefs in his "arthritis" example, it is now time to take stock of the situation. Let us assume, for the sake of taking stock, that Loar is correct in his claims about the psychological content of the beliefs involved in Burge's "arthritis" example. That is, in the actual situation, the psychological content of Bert's arthritis belief is distinct from the doctor's arthritis belief; also, that the psychological content of actual-Bert's arthritis belief is the same as the psychological content of counterfactual-Bert's arthritis belief. This difference or sameness of psychological content of the arthritis beliefs is evidenced by the difference or sameness of the inferences and behaviour that these beliefs elicit in the people concerned, i.e., how the beliefs interact conceptually with the subject's other beliefs and attitudes in the psychological explanation of the subject's behaviour.

Of course, all of this can be accepted whilst still accepting Burge's conclusions about the belief ascriptions involved in the example (which Loar himself does accept).

Namely, in the actual situation, one can truly ascribe the belief "Bert has arthritis in his ankles" to both Bert and the doctor, whilst in the counterfactual situation, one cannot truly ascribe the belief "Bert has arthritis in his ankles" to Bert. According to

Loar, this would be a case of the social content that de dicto that-clause belief ascriptions contain, being sensitive to changes in the linguistic or social environment of the subject who is being ascribed the belief. This then is Loar's content-dualist position, the narrow psychological content of a subject's belief staying the same between actual and counterfactual situations, while the wide social content of the belief differs.

Overall, Loar's content-dualist position does seem fairly plausible, and it has been a position which I have generally been trying to defend. However, I have a niggling feeling that there are certain aspects to Loar's position which are somewhat problematic. As I see it, the biggest problem is caused by Loar's contention that that-clauses do not generally capture the psychological content of a subject's belief, as that content is individuated in commonsense psychological explanation. This claim will immediately seem strange to many, as it is generally assumed that "commonsense psychological explanation" just is the use of a that-clause which has a certain "content" and which is preceded by a particular psychological verb, that helps to explicate the behaviour of a subject,<sup>1</sup> e.g., that Bert took an umbrella with him when he went out, is explained by his belief that it was going to rain later and his desire not to get wet.

If Loar is advocating that that-clauses are generally out of bounds to describing or specifying the narrow psychological content of beliefs and other attitudes, then it seems that his options for giving commonsense explanations of behaviour involving this psychological content have been seriously curtailed. This seems to be the reason why Loar is reduced to just claiming that there are various ways to identify when a

belief has a narrow content. For example, to note the contexts in which that-clauses are used to describe beliefs, so that "...different narrow contents are implied by "Paul believes that cats have tails" in the "chat" context and in the "cat" context" (Loar, 1988a, p.109). Another way is to note the words that people use to describe their beliefs, in order to gauge if there is a difference of conceptual role in the subject's use of a word, in contrast to its usual social meaning, e.g., when Bert believes "that he has arthritis in his ankles", but goes on to say that he believes "that he has arthritis in his thigh" as well, this indicates that his conception of arthritis has a narrow content (Loar, 1988a, p.109).

Indeed, Loar goes so far as to claim that perhaps we don't actually need specifications of narrow content, but that we must still recognise that it does exist, as is shown with the conceptual role-individuation of beliefs in the Pierre, Paul and Bert examples (Loar, 1988a, p.109). However, it might plausibly be claimed here that if narrow psychological content is going to play a *general* explanatory role, rather than just doing so on some specific occasions, then surely it must be specifiable by some means. Moreover, it seems to me that the only plausible means available for a general specification of psychological content is through the use of that-clauses. However, narrow content fans need not fear, even though I am considering "disinvesting" in Loar's views on that-clauses, I am not giving up on narrow content or Loar altogether, as will be seen in a short while.

In order to try to explain my position on this issue, let us return to Burge's arthritis example, to the actual situation of Bert and the doctor. That the conceptual roles of Bert's and the doctor's arthritis beliefs are different seems clear enough, indeed, this

is a point that it seems Burge was quite willing to concede. Burge's claim was that despite the difference of conceptual roles, the two beliefs would be considered the same as far as commonsense ascription individuated the beliefs, i.e., as shown in oblique univocal that-clauses. Loar basically agreed with Burge on this point, but took issue with him when he made the further claim that the arthritis beliefs of Bert and the doctor also had the same content as that was individuated in commonsense psychological explanation. Loar believed that the arthritis beliefs of Bert had a different psychological content to the arthritis beliefs of the doctor, despite their having the same de dicto that-clause descriptions (or social content), and that this was evidenced by the inferences and behaviour of Bert.

In examining Loar's thinking on the above point, I have found it useful to utilise so-called subjective and objective points of view (re Thomas Nagel) when ascribing beliefs to a subject to explain his or her behaviour. It seems to me that when ascribing beliefs to a subject, one can use the subjective viewpoint in doing this by either imagining that one has asked the subject to describe the reasoning that led to his/her behaviour, or one could use the disquotation principle, where the subject assents to or dissents from certain statements made by the ascriber. When beliefs are ascribed from an objective viewpoint, the ascriber seems to put more emphasis on facts in the social environment of the subject, as if the ascriber is ascribing to the subject the beliefs that he/she should have, given the state of affairs in their current environment (this difference between subjective and objective viewpoints when making belief ascriptions will come to the fore when Bert's counterfactual situation is examined). When Loar talks of commonsense psychological explanation, I think he means as this

is perceived from the subject's own viewpoint, what the subject him-/herself takes to be the case in explaining his/her inferences and behaviour.<sup>2</sup>

However, I think it is here that the puzzling nature of Loar's claim that that-clauses are unable to specify the narrow psychological content of a subject's beliefs quickly becomes evident. Let us consider Bert's belief that he has arthritis in his thigh. It is because of this belief, and the idiosyncratic understanding of the concept of *arthritis* that it seems to entail, that the conceptual role of Bert's arthritis beliefs are considered different from the conceptual role of the doctor's arthritis beliefs. Now, imagine that Bert has a number of hypodermics containing cortisone which he understands will ease the pain of arthritis. We observe him injecting himself in the ankles, but then he injects himself in his thigh also. If the subjective viewpoint of Bert is now used in giving a psychological explanation of why he injected himself in the thigh, then it seems fairly clear that it will go something like this: I (Bert) had the belief that the cortisone would ease the pain of my arthritis, I believed that I had arthritis in my thigh and so injected the cortisone into my thigh.

The upshot of this is that it seems that the psychological content of Bert's beliefs have been captured using that-clauses.<sup>3</sup> Moreover, the that-clauses also seem to effectively describe, from the subjective viewpoint, how the psychological content of Bert's beliefs interact with his other beliefs to produce his behaviour. Indeed, it seems that in this case the ascription of beliefs to Bert from the subjective viewpoint, are the same as the ascription of beliefs from the objective viewpoint, i.e., from this latter viewpoint it would be true that Bert was having arthritis beliefs (as it would also be true that the doctor was having arthritis beliefs) .



What would Loar say about this situation? Perhaps he would say that even though it appears that Bert and the doctor have the same arthritis beliefs, in the sense that their beliefs have the same social content, the psychological content of Bert's arthritis beliefs would be different, because he has a different concept of *arthritis* than the doctor does. The result of this would be that even though it appeared that we could describe Bert's psychological reasoning with that-clause ascriptions containing the word "arthritis", the content that the word would have would be its social content, i.e., a disease of the joints only, which would not be the content that it had as understood by Bert, otherwise he would not believe that he had arthritis in his thigh.

This is of course roughly the position that I have been trying to previously defend, however, it now seems to be a rather vulnerable position, especially if Loar is hoping to argue that beliefs in general, all have a narrow psychological content. Presumably, once Bert correctly understands the concept of *arthritis*, then Loar would contend that the psychological content of his arthritis beliefs would be the same as the social content. However, one can imagine what the parsimonious reply to Loar's contention would then be: instead of assuming that Bert's arthritis beliefs have an unspecifiable, narrow, psychological content and a specifiable, wide, social content, it seems much more plausible just to assume that the beliefs only have a specifiable, wide, social content, which at first, Bert misunderstands.

However, it is important to reiterate that this rather awkward situation is only reached if we go along with Loar's presupposition that that-clauses do not generally capture the narrow psychological content of beliefs. In claiming this, Loar has effectively

surrendered the use of that-clause belief ascriptions to the externalists, which makes things difficult, as it seems that it is only by the use of ordinary that-clauses that some kind of commonsense psychological explanation can be given, either from the subjective or objective point of view. Therefore, it seems to me that if some type of narrow content is to sound feasible, then the use of that-clauses must be brought back on to the side of the internalist, i.e., so that-clauses can capture narrow content. This will be the objective of the next section.

## **2.1 An Attempt to Specify the Narrow Psychological Content of the Belief Ascriptions in Burge's "Arthritis" Thought Experiment**

This section will be concerned with trying to find that-clause belief ascriptions that can specify some sort of conceptual role narrow content in Burge's "arthritis" thought experiment. To facilitate this, I will continue to use the subjective and objective distinction between belief ascriptions; also, de dicto and de re belief ascriptions will be examined, in order to determine if one of these modes of ascription can be utilised to capture the narrow content of a belief.

### **2.1.1 *The Actual Situation***

I will now examine the arthritis beliefs of a subject, Bert, in the actual and counterfactual situations of Burge's thought experiment.<sup>4</sup> Therefore, let us first consider Bert's belief that he has arthritis in his thigh in the actual world. To begin

with, how does one ascertain whether or not to ascribe this particular belief to Bert? Firstly, one could listen to Bert's utterances, and if he uttered "I have arthritis in my thigh", then this would be a good reason to then ascribe to Bert the belief that he had arthritis in his thigh. One could also use the disquotation principle of belief ascription, exploiting Bert's sincere assent to and dissent from particular statements which are put to him. For example, the ascriber could say to Bert something like "you believe that you have arthritis in your thigh", a statement which we would assume that Bert will assent to.

The upshot of this would be that we could plausibly ascribe to Bert the belief that he had arthritis in his thigh. As previously mentioned, these two methods of ascribing beliefs do justice to the subjective first-person perspective of the subject, how the subject conceives of things, that is, how things seem to Bert, given the particular context he is located in. What I take to be the objective third-person perspective on ascribing beliefs will also probably use the above methods; however, it seems to me that the objective third-person perspective additionally includes factors in the social environment of the subject, in determining the truth or falsity of the subject's conception, and so deciding what belief can truly be ascribed to the subject.<sup>5</sup> In the actual world, it seems that the belief ascribed to Bert from the objective viewpoint will be the same as the one ascribed from the subjective viewpoint, that is:

(B1) Bert believes that he has arthritis in his thigh.

However, it seems to me that it is the following question that is important in connection with establishing the possibility of some sort of narrow content shared by

Bert's actual/counterfactual arthritis beliefs: is (B1) a *de dicto* or a *de re* belief? This distinction is a controversial one, however, what may be termed the "orthodox position" is that:

Belief *de dicto* is belief that a certain *dictum* (or proposition) is true, whereas belief *de re* is belief about a particular *res* (or thing) that it has a certain property.<sup>6</sup>

The *de dicto* and *de re* distinction of beliefs involves many more aspects than are mentioned above.<sup>7</sup> However, rather than just listing them here, I will now examine how Burge himself makes the distinction in relation to his arthritis thought experiment, and hopefully the various aspects will emerge as we progress.

#### Burge's De Dicto/De Re Distinction

Burge holds that (B1) is a *de dicto* or oblique belief, mainly because the general term "arthritis" is occurring obliquely in the that-clause which appears in (B1). He believes that it is expressions occurring obliquely "...within content clauses that primarily do the job of providing the content of mental states or events, and in characterizing the person" (Burge, 1979, p.76). For Burge, an oblique expression within a content-clause is an expression which need not be intersubstitutable with "extensionally equivalent expressions in such a way as to maintain the truth value of the containing sentence" (Burge, 1979, p.76).

For example, even though "arthritis" is defined as an "inflammation of the joints", just because Bert believes that he has arthritis in his thigh, this need not entail that he

would also believe that he had an inflammation of the joints in his thigh. For Bert, the extensionally equivalent expressions have different cognitive significance. Moreover, it is a commonly accepted view that the content of a mental state or propositional attitude is what is contained in the content clause or that-clause of the particular state or attitude in question (Burge, 1979, p.75). Therefore, in (B1) above, the content of Bert's belief would be the expression "that he has arthritis in his thigh". So, it would be this oblique expression that would describe Bert's subjective viewpoint on his situation, that is, how he conceives things as being in the world.

Importantly, for Burge, it is *de dicto* beliefs with that-clauses containing oblique expressions, which essentially capture the content and subjective character of a person's intentional states. He contrasts this with "non-oblique occurrences of expressions in content clauses" (Burge, 1979, p.76), by which I take Burge to mean *de re* beliefs. A *de re* belief is considered to be a belief *of* an object, which involves the subject having the belief being in an indexical or contextual relation to the object in question (Burge, 1979, p.86; 1982, p.118). According to Burge, the non-oblique expression in a *de re* belief would not form a part of, or characterise, the content of that belief. If I have understood Burge correctly, then the belief ascription:

(B2) Bert believes that the disease known as arthritis is what he has in his thigh

would be considered by him to be a *de re* belief ascription. With (B2), the occurrence of the expression "arthritis" in the that-clause would be non-oblique, as (according to Burge) the ascriber may believe that "no distinction... would be lost" if an extensionally equivalent expression were substituted for "arthritis", e.g., one could

use the expression “inflammation of the joints” instead, even if Bert himself was unaware of the equivalence of these expressions (Burge, 1979, p.76).<sup>8</sup> Moreover, the expression “arthritis” in (B2) would not be part of the content of Bert’s belief; rather, he would believe his belief content *of* arthritis (Burge, 1979, p.76; p.86). Burge believes that the non-oblique occurrence of the term “arthritis” simply facilitates an indexical or contextual relation between the disease arthritis and the contents of Bert’s belief. The upshot of this would be that *de re* beliefs essentially concern how Bert’s belief-content is related to the world, rather than essentially constituting and characterising his belief-content (Burge, 1979, p.86). To recap, this is basically the reason that Burge argues that the beliefs involved in his “arthritis” thought experiment are not *de re*, but are *de dicto* ones, containing obliquely occurring expressions in that-clauses which serve to make up the content of the beliefs.<sup>9</sup>

### The Possibility of Subjective and Objective *De Re* Beliefs?

Having set out what I take to be Burge’s position on the *de dicto/de re* distinction in relation to his “arthritis” thought experiment, it is now time to consider some of the issues that have been raised. Contrary to Burge, McKinsey (1993) claims that the belief ascriptions involved in the arthritis example can be given a *de re* reading, as well as a *de dicto* one. He states that on a *de re* reading of the beliefs:

...the general term is given large scope and the believer is said to have a belief that is referentially about the type meant by the general term. In this case, the belief ascription can be true even though the believer does not have an adequate understanding of the concept expressed by the term, and it is in this sense that it is true in Burge’s case that [Bert] believes that he has arthritis in his thigh (McKinsey, 1993, p.334).

What could be said to counter McKinsey's above claim about Bert's belief that he has arthritis in his thigh? Presumably one could again point out that the term "arthritis" is occurring obliquely in the that-clause of Bert's de dicto belief, and is forming the content and character of his belief, something which a non-obliquely occurring "arthritis" in a de re belief would fail to do (Burge, 1979, 76; p.87).

However, it seems to me that the issue is not as clear cut as it appears. For instance, bearing in mind the earlier subjective/objective distinction that was made in the readings of Bert's arthritis belief, the following possibility presents itself. What if one considers de re readings of Bert's belief that he has arthritis in his thigh from subjective and objective viewpoints? From the objective point of view, the ascriber's use of the term "arthritis" in the that-clause of the belief ascription could be correctly described as being non-oblique, as it does not matter to the ascriber whether "arthritis" or the extensionally equivalent expression "inflammation of the joints" is used in the ascription. That is, on an objective de re ascription, the content of Bert's belief is being used to refer to the disease known as arthritis in Bert's linguistic community. Bert himself does not know the concept that is expressed by the term "arthritis", but this does not matter from an objective viewpoint, as the ascriber does know the concept that is expressed by "arthritis", and is tacitly referring to it using Bert's belief.

However, what if Bert has a de re belief from his subjective viewpoint? By this I mean that the term "arthritis" could still be used by Bert to refer to the disease called arthritis in his linguistic community, but it would be used in such a way that it did

form and characterise the content of his belief. The term “arthritis” would be oblique, as if the extensionally equivalent expression “inflammation of the joints” was substituted for it in Bert’s belief, the truth value of the containing sentence would change, i.e. he would not believe that he had inflammation of the joints in his thigh. However, despite the obliqueness of the term “arthritis” and the characterising of Bert’s mental state by this term, it seems to me that “arthritis” could still have a *de re* meaning to Bert himself, something along the lines of referring to the disease called arthritis in my surrounding social and linguistic community. Therefore, from the subjective viewpoint of Bert himself, we seem to have an oblique expression in a that-clause that forms and characterises the content of his belief, but that nevertheless has a *de re* meaning to Bert.

At first glance, the talk of a subjective *de re* belief, as opposed to an objective *de re* belief, sounds a promising route to locating some narrow content that is contained in Bert’s actual arthritis belief. However, on reflection, I think that this strategy is misguided. The reason for this involves the way that *de re* beliefs are individuated, i.e., what the method is for saying that certain *de re* beliefs are either the same or different. The point can be illustrated using Quine’s (1956) notational device (with some modifications by Stich (1986)) for illustrating the structure of a “relational” or *de re* belief. Quine believed that a *de re* belief involved a three-place relation between a person, an open sentence and an object (Stich, 1986, p.120). For example, on a “relational”(r) or *de re* reading, Bert’s belief that he has arthritis in his thigh would be:

Bert believes(r) (‘that he has ... in his thigh’, <arthritis>).



It can be seen that with this relational reading of the belief, the person is Bert, the open sentence is ‘that he has ... in his thigh’, and the object is the disease called “arthritis”. The important point for the issue at hand is that on a relational de re reading of a belief about a particular object, the belief is individuated by that object. So in the case of Bert’s de re arthritis belief, whatever expression is put in the place of the ellipsis in the open sentence, the belief will still be individuated by the disease arthritis. On a de re reading, the non-oblique or transparent occurrence of the expression “arthritis”, means that it has a specific referential function, i.e., to refer to the disease called arthritis in that context. The upshot of this is that whether we consider Bert’s actual belief that he has arthritis in his thigh as a subjective or objective de re belief, both belief-readings are going to be *of*, or have a contextual relationship with, exactly the same arthritis concept, i.e., a rheumatoid inflammation of the joints.

Therefore, even though Bert may not know what the meaning or concept of arthritis is in the actual world, on either a subjective or objective de re reading of his arthritis belief, the expression “arthritis” can still be intersubstitutable with the extensionally equivalent “inflammation of the joints”, as the latter is the conceptual referent of the expression “arthritis”. In effect, when a de re arthritis belief is ascribed to Bert, the ascriber is using Bert’s belief to point to the disease called “arthritis”; or as Burge would put it, Bert’s de re arthritis belief is illustrating the contextual relationship between his belief-content (which would be ‘that he has ... in his thigh’) and the disease arthritis. Therefore, the idea of subjective and objective de re readings of Bert’s actual arthritis belief will not be able to produce a narrow meaning of “arthritis”.

## The Possibility of Subjective and Objective De Dicto Beliefs?

However, what if we consider the possibility of subjective and objective de dicto readings of Bert's belief that he has arthritis in his thigh? Before we do this, let us first examine how exactly an oblique or opaque occurrence of the term "arthritis" is supposed to constitute and characterise the content of the de dicto belief that contains it. In the case of Bert, the obliqueness of the term "arthritis" means that it is true that he does believe that he has arthritis in his thigh, but it need not be true that he believes he has the extensionally equivalent "inflammation of the joints" in his thigh. But what does this distinction amount to with regards to providing and characterising the content of Bert's belief? It seems to me that all that comes out of this is that Bert believes he has a disease called arthritis in his thigh, the same disease which he also believes he has in his ankles and wrists. However, McKinsey proposes that with a de dicto belief:

...the general term is given small scope, and as a result the person is said to have a belief one of whose elements is semantically analogous to the general term in question. In this case, the general term's meaning is invoked to characterize the believer's way of thinking of the type meant by the term, and thus full understanding of the concept expressed is presupposed (McKinsey, 1993, p.334).

It is for this reason that McKinsey claims that it would be incorrect for Burge to say that Bert truly believes he has arthritis in his thigh on a de dicto reading, because this would amount to saying that Bert believes he has inflammation of the joints in his thigh, something which he probably would not believe (McKinsey, 1993, p.334).

However, is McKinsey correct in claiming that a term in a de dicto belief constitutes and characterises the content of that belief by providing the conventional meaning that

the term has in that linguistic community? Interestingly enough, it seems that support for this view may come from Burge himself, where he states that:

An attitude *content* is the semantical value associated with oblique occurrences of expressions in attributions of propositional attitudes... Thus the content is, roughly speaking, the conceptual aspect of what a person believes or thinks (Burge, 1982, p.119).

It is tempting here to conclude that Burge is roughly saying the same thing as McKinsey did above. That is, the content that an oblique expression in a that-clause provides to the attitude attributed, is the conventional meaning or concept associated with that expression. This view also seems to be supported by Loar (1988a), while discussing the issue of whether “believes that there is a cat in the next room” could be truly ascribable to both Helen Keller and a sighted person, Loar declares that:

For if each were sincerely to assert “there is a cat in the next room” their words conventionally would *mean* the same; and so by ordinary criteria the belief ascription would be true of each on an *oblique* or *de dicto* univocal reading. (The sameness of belief ascription is not then merely a function of a common *de re* reference to the kind *cat*.) (Loar, 1988a, p.100).

It can be seen here that Loar is saying that a *de dicto* reading of a belief ascription involves the conventional meaning of the words contained in the ascription. This is in contrast to the referential element that would be involved with a *de re* reading of the words in the belief ascription. It is for this reason that Loar agrees with Burge that in the actual context, Bert’s belief that he has arthritis in his ankles and the doctor’s belief that Bert has arthritis in his ankles are the same on a univocal *de dicto* reading (Loar, 1988a, p.100). Of course, it was at this point that Loar proposed that the sameness of *de dicto* arthritis beliefs of Bert and the doctor need not imply sameness of psychological content of those beliefs. As the beliefs were individuated by

commonsense psychological explanation, Bert's arthritis belief seemed to be distinct from the doctor's arthritis belief. Loar put this distinctness of beliefs down to the different conceptual role that Bert's arthritis belief had in his cognitive life, which was evidenced by his resultant behaviour.

Therefore, Bert had the belief that he had arthritis in his ankles while he also had the belief that he had arthritis in his thigh, that is, before he was corrected by the doctor and told what the actual definition of arthritis was. This seems to show that while the oblique occurrence of the term "arthritis" in the content-clause of the doctor's belief about Bert could be said to be expressing its conventional concept or meaning, the oblique occurrence of the term "arthritis" in the content-clause of Bert's belief could *not* be said to be expressing its conventional concept or meaning. Therefore, it seems to follow that while we could truly ascribe the belief "that Bert had arthritis in his ankles" to the doctor on a *de dicto* reading, we could *not* truly ascribe to Bert the belief "that he had arthritis in his ankles" on a *de dicto* reading. It was for this reason that McKinsey proposed above that Bert could truly believe he has arthritis in his thigh only on a *de re* reading of his belief, as the term "arthritis" would be referring to the conventional concept meant by the term. This *de re* reading would not presuppose that Bert had a full understanding of the concept of arthritis (something which McKinsey thought that a *de dicto* reading of the belief did presuppose).

McKinsey's suggestion for a *de re* reading of Bert's arthritis belief does seem plausible, but his strategy for finding some narrow content in Bert's belief is different from the one I am adopting here. I am roughly following Loar's line that there is an apparent bifurcation of content that is not recognised on the orthodox understanding

of Bert's actual arthritis beliefs. As previously shown, considering Bert's belief that he has arthritis in his thigh on a de re reading, does not result in the required content-bifurcation. I am now considering Bert's belief on a de dicto reading, which was what Loar did himself, and which resulted in what he called the psychological content and social content of a person's de dicto belief. This result is roughly analogous with the subjective and objective distinction that I previously made in Bert's arthritis belief, so it seems that we may be on the right track. Again using Quine's notation (along with Stich's modifications), a "notional"(n) (de dicto) reading of Bert's belief that he has arthritis in his thigh would be:

Bert believes(n) 'that he has arthritis in his thigh'.

It can be seen that a de dicto belief is considered to be a two-place relation between a person and a sentence. In this case, Bert is the person and the sentence is 'that he has arthritis in his thigh'. In contrast to the de re reading, the sentence is not open and there is no object which closes it, and which consequently contributes to the individuation of the belief. The oblique occurrence of the term "arthritis" in Bert's de dicto belief is considered to individuate the belief and so provide and characterise the content of the belief from his perspective. Because there is the possibility that the term "arthritis" in Bert's belief may not be substituted with an extensionally equivalent expression and the truth value of the belief stay the same, it is usually said that the term in question does not have a specific referential function. (Although of course the oblique term in a de dicto belief does have a general referential function, this latter function does not individuate the belief to the extent that the referential function of a de re belief would – I think this makes sense).

It seems to me that a de dicto reading of Bert's belief that he has arthritis in his thigh is a more promising option for making a subjective/objective bifurcation of belief content. For instance, one could say that an objective de dicto reading of Bert's arthritis belief could supply the conventional meaning and conceptual content to the term "arthritis". That is, on an objective de dicto reading, Bert would believe that he had an inflammation of the joints in his thigh, which would be a false belief, but it could still be said that Bert truly believed this. The objective de dicto reading of Bert's arthritis belief would be equivalent to Loar's social content of Bert's de dicto belief.

However, it seems clear that the objective de dicto reading of Bert's arthritis belief is not very satisfactory, or indeed very plausible. Of course, it might seem satisfactory and plausible from the objective third-person perspective of the ascriber, who is following the semantical rules of de dicto belief ascription that McKinsey, Burge and Loar all seem to support. From this perspective, the ascriber gives the term "arthritis" its conventional meaning and this is supposed to provide and characterise the content of Bert's arthritis belief. However, surely this is not satisfactory or plausible when one takes the subjective first-person perspective of Bert himself. After all, if a de dicto belief is one that involves a person believing that a particular proposition is true, then if Bert truly believes that he has arthritis in his thigh, it seems plausible to claim that whatever meaning Bert puts to the term "arthritis", it is not going to be something like "inflammation of the joints", at least if we consider Bert to be rational and to know that there is not a joint in his thigh.

Indeed, the orthodox procedure for dealing with the meanings/concepts of oblique terms in de dicto beliefs seems a little odd, at least when those meanings/concepts are misunderstood or only partially understood. The obliqueness of the relevant term is supposed to recognise that the believer may attach a different semantical value to that term than it conventionally has, as there is the possibility that the term may not be substituted with a co-extensional term and keep the truth value of the belief the same. This sounds great for recognising that the believer himself may have a different conception of a term's meaning to what the conventional meaning is. However, what happens next is that from the third-person perspective, the ascriber ascribes a de dicto belief to the believer which has the conventional meaning attached to the relevant term in the belief. Moreover, from the third-person perspective, this objective meaning that is attached to the term in question is to provide and characterise the content of that term for the believer. However, this whole procedure is geared towards the provision and characterisation of belief-content from the objective third-person perspective of the ascriber himself. If the believer does know the conventional meaning/concept of the term in question, then this procedure is successful, but if he does not, then it seems to me that his subjective idiosyncratic understanding of the term is not really specified or characterised by the orthodox de dicto belief ascribed to him.

Therefore, even though the general term in a de dicto belief ascription is oblique rather than non-oblique, if I have understood things correctly (viz., the previous claims of McKinsey, Burge and Loar on de dicto belief ascriptions), the ascriber of the de dicto belief will still apply the conventional meaning or concept to the general term in question, in order to provide and characterise the content for that belief

ascription.<sup>10</sup> Indeed, perhaps one could say that from the third-person perspective of the ascriber, a de dicto belief containing an oblique term behaves like a de re belief which contains that term, as the ascriber is using the oblique term in the ascription to specifically refer or point to, the relevant object or concept in question.

Nevertheless, it is roughly at this point that Loar proposes his psychological content with regards to Bert's arthritis belief, which would be a narrow content that evidenced itself in the conceptual role that "arthritis" played in the interactions with Bert's other thoughts and the resultant behaviour that this produced. It will be recalled that the problem with Loar's psychological content of a belief was that it was not easy to specify what it was, and the concept also seemed questionable, as Bert would still use the term "arthritis" in commonsense psychological explanations of his behaviour. Therefore, what it seems we need from a subjective de dicto reading of Bert's arthritis belief is a narrow content similar to Loar's psychological content, but which is specifiable and enables us to use the term "arthritis" in psychological explanations of Bert's behaviour. Perhaps we could use Quine's and Stich's notation to help separate out the objective and subjective meanings of the term "arthritis" in Bert's de dicto belief. For instance, an objective de dicto reading of Bert's belief would look like:

Bert believes(on) 'that he has arthritis in his thigh'.

where the "(on)" could stand for "objective-notional" and the semantic content that the oblique term "arthritis" has could be the conventional meaning, which in the actual context, is a "rheumatoid inflammation of the joints". As stated above, it is the ascriber, from the third-person perspective, who would ascribe this objective de dicto



belief to Bert, with the semantic value of the oblique “arthritis” having its objective social meaning. With regards to a subjective de dicto reading of Bert’s belief, one could have the following:

Bert believes(sn) ‘that he has arthritis in his thigh’.

where the “(sn)” could stand for “subjective-notional” and the semantic content that the oblique term “arthritis” has could be the subjective meaning which Bert attaches to the term in the actual context. From the third-person perspective, the ascriber could still ascribe the subjective de dicto belief to Bert using the oblique term “arthritis”, but he would not give the term its conventional meaning.<sup>11</sup>

Therefore, the question now is what semantic content “arthritis” would have in Bert’s subjective de dicto belief? It is here that things get rather problematic, as of course Bert could basically attach *any* meaning to the term “arthritis”, as that term could now be claimed to be a private language term, with its own private meaning for Bert.

Moreover, any restrictions on the meanings that Bert can attach to “arthritis” will seem to be arbitrary. I suppose that something similar to a “principle of charity” or “principle of humanity” for belief ascription could be utilised here. A principle of charity is where an interpreter assumes that a subject has certain beliefs, given the situation they are in, and that the beliefs of the subject will be mostly true, and not be totally implausible (Braddon-Mitchell/Jackson, 1996, p.147). The principle of humanity is similar, but here the interpreter assumes that the beliefs a subject has would be similar to the beliefs that the interpreter would have, if they were in the subject’s position (Braddon-Mitchell/Jackson, 1996, p.147). The upshot of both these

principles seems to be that the subject who is being ascribed beliefs is assumed to be a sane and rational individual, just like the interpreter.

So perhaps we can assume that Bert is a sane, rational and averagely intelligent person, who will not attach an outlandish meaning to “arthritis”, e.g., he would not believe that “arthritis” is a type of grass, etc. This still seems to have an arbitrary air about it, nevertheless I am going to push on, as in the actual context, it seems to me that the most plausible meaning that Bert would attach to “arthritis” would be something like “the disease called arthritis that I also have in my ankles and wrists, that other people have in their ankles and wrists, that my father had in his shoulders, etc.” Therefore, from the first-person perspective of Bert, the subjective content of “arthritis” is not going to be a specific individual concept as such, but more of a relational concept, that his arthritis is a disease that is the same as the disease called arthritis in his present social environment. Now, this relational concept that I am attributing to Bert’s “arthritis”, sounds very much like a *de re* component to Bert’s belief, in that his term “arthritis” is specifically referring to the disease called “arthritis” in the actual context. The upshot of this would be that the semantic content of Bert’s “arthritis” would be its social meaning, i.e., a rheumatoid inflammation of the joints (which is also what its objective *de dicto* meaning would be).

Of course, this is just what we don’t want, as we will be back to the situation of attributing an implausible semantic content to the term “arthritis” in Bert’s belief. The semantic content of “arthritis” in Bert’s belief, on a subjective *de dicto* reading, is just going to be something like the rheumatoid disease in my (Bert’s) social environment, full stop. That description is going to be the concept of “arthritis” in the

actual context, as far as Bert is concerned. Bert's *arthritis* concept is not going to go one step further, and include what arthritis actually is, i.e., a rheumatoid inflammation of the joints. However, despite this, it is important to point out that from Bert's subjective viewpoint, the content of his belief is still being constituted and characterised by the term "arthritis". This is due to the oblique occurrence of "arthritis" in Bert's belief, in that if the term "arthritis" was substituted by a co-extensional expression, such as "inflammation of the joints", the truth value of the containing sentence would change. That is, Bert would not truly believe that he had inflammation of the joints in his thigh, so it seems that the term "arthritis" has a distinct cognitive value to Bert. Therefore, it seems that the narrow content of Bert's arthritis belief is able to be specified, and we can still use the term "arthritis" to attribute a narrow belief content to Bert, as long as it is understood that the belief has a subjective-notional de dicto reading.

Moreover, Bert's subjective de dicto arthritis belief is an instance of a recurring theme in the narrow individuation of thoughts and beliefs. The belief is narrowly individuated by a definite description of the object or concept concerned, but the description does not fix the actual object or concept concerned. The upshot of this is that in another context, one could say that the subject's narrow belief content is the same, as there is an object or concept that satisfied the definite description, even though that object/concept may well be different to the object/concept in the original context. In contrast to this, the wide individuation of beliefs will involve the actual object/concept in the belief content, and not just leave the determination of the object/concept to a definite description. This means that in another context where the

object/concept is different, then the belief content will also be different, even though the object/concept will satisfy the same definite description in that new context.<sup>12</sup>

Nevertheless, it should be pointed out that Burge himself is not concerned whether the beliefs in his arthritis example are viewed as being de re or de dicto, as he claims that the important point is “...that the social differences between the actual and counterfactual situations affect the *content* of the subject’s attitudes...the difference affects standard cases of obliquely occurring, cognitive-content-conveying expressions in content clauses” (Burge, 1979, pp.86-87). Therefore, to see whether Burge is correct in this claim, the counterfactual situation of his thought experiment must be considered, and this is what we go to next.

### 2.1.2 *The Counterfactual Situation*

In the counterfactual situation, Bert is considered to be physically identical to the Bert in the actual situation. That is, actual-Bert and counterfactual-Bert are identical with regards to physiological, sensory, phenomenal, dispositional, behavioural, functional etc. properties and experience (non-intentionally described) (Burge, 1979, p.78). The only difference between the actual and counterfactual situations is in the “social environment” of the latter. Burge supposes that in the counterfactual situation, the word “arthritis” refers to various rheumatoid ailments that can occur outside the joints, as well as inside the joints, in contrast to “arthritis” in the actual world, which only refers to rheumatoid ailments inside the joints (Burge, 1979, p.79). The result of this difference is that the ailment Bert has in his thigh may well fall under the

definition of “arthritis” in his counterfactual social community, while it did not fall under the definition of “arthritis” in his actual social community. Burge then gives his interpretation of the counterfactual situation:

In the counterfactual situation, the patient lacks some – probably *all* – of the attitudes commonly attributed with content clauses containing ‘arthritis’ in oblique occurrence. He lacks the occurrent thoughts or beliefs that he has arthritis in the thigh, that he has had arthritis for years, that stiffening joints and various sorts of aches are symptoms of arthritis, that his father had arthritis, and so on (Burge, 1979, p.78).

I think that Burge’s reasoning in reaching the above conclusion goes something like the following. Because the term “arthritis” in the actual world denotes, or just means, “inflammation of the joints”, then for us (in the actual world) “arthritis” is necessarily going to denote or mean “inflammation of the joints” in all possible worlds.

Therefore, when a content clause containing an oblique occurrence of “arthritis” is ascribed to counterfactual-Bert, the denotation or meaning of “arthritis” is going to be what it is in the actual world. However, according to Burge, it would be incorrect to ascribe to counterfactual-Bert any content clauses containing the actual term “arthritis”, as this term has a different denotation or meaning in the counterfactual world, i.e., in the counterfactual world, the term “arthritis” denotes or means “inflammation of the joints and bones” (or something like that).<sup>13</sup>

The upshot of this difference is that counterfactual-Bert cannot correctly be ascribed a content clause, or attributed an attitude, which contains an oblique occurrence of the term “arthritis”. From this conclusion, Burge takes the further step of claiming that the content of actual-Bert’s arthritis thoughts will be different from the content of counterfactual-Bert’s arthritis thoughts. This step follows naturally from Burge’s

supposition that if counterpart expressions in ascribed content clauses have different extensions, then the attitudes attributed using those content clauses will have a different content (Burge, 1979, p.79). Moreover, this difference in the content of the mental states of the subject will have been produced only by a difference in the counterfactual social environment of that subject.

### A Possibility of Narrow Belief Content?

It is interesting to wonder whether any sort of narrow content can be discerned in the actual and counterfactual situations of Burge's thought experiment, i.e., is there some "arthritis" content that actual-Bert and counterfactual-Bert share, despite the differences in their respective social environments. Therefore, consider the belief which could be expressed by:

Counterfactual-Bert believes that he has arthritis in his thigh.

On either a subjective or objective de re reading, it seems clear that the semantic content of counterfactual-Bert's "arthritis" is going to be different to the semantic content of actual-Bert's "arthritis". It will be recalled that the ascription of a de re belief is considered to involve a three-place relation between a person, an open sentence and an object, so on a de re reading of the above belief, we would have

Counterfactual-Bert believes(r) ('that he has ... in his thigh', <arthritis>).

The important point is that the above belief is going to be individuated by the object, which in this case is the counterfactual concept of arthritis. As the concept of arthritis in the counterfactual context is something like “an inflammation of the joints and bones” then it is this semantic value that the non-oblique occurrence of “arthritis” in the open sentence is going to have, whether counterfactual-Bert realises this or not. As before, the non-oblique occurrence of “arthritis” in the belief ascription is a method which can be used by the ascriber to point directly to the extension of that term in that particular context. The upshot of this is that actual-Bert’s subjective/objective de re belief is going to have an *arthritis* concept like “inflammation of the joints”, whilst counterfactual-Bert’s subjective/objective de re belief is going to have an *arthritis* concept like “inflammation of the joints and bones”. Therefore, this difference in extensional concepts of the respective Berts’ “arthritis” terms means that the semantic content of the “arthritis” terms are also going to be different, so it seems there is no possibility of a shared narrow content on a de re reading of the Berts’ beliefs.

However, what is the situation with regards to a subjective and objective de dicto reading of counterfactual-Bert’s arthritis belief? An objective de dicto reading of the belief would look like

Counterfactual-Bert believes(on) ‘that he has arthritis in his thigh’.

Once again a de dicto belief is a two-place relation between a person and a sentence, with the objective-notional version giving the oblique “arthritis” the semantic content of whatever is the conventional meaning of arthritis in that context. The result of this

is that in the counterfactual context, the semantic content of the term “arthritis” will be “an inflammation of the joints and bones”. Therefore, on an objective de dicto reading, the semantic content of counterfactual-Bert’s arthritis belief will be different to the semantic content of actual-Bert’s arthritis belief. Moreover, this difference of belief contents has come about only through a difference in the social environment of counterfactual-Bert. This stems from Burge’s claim that if there is a difference in extension between two counterpart terms, which make up the content of two mental states, then the contents of those two mental states would be considered different (Burge, 1979, p.75), and, eo ipso, the two mental states themselves would be considered distinct. This certainly seems to be the case with the “arthritis” terms which actual-Bert and counterfactual-Bert are using in their respective contexts. In the actual context, the extension of the term “arthritis” is “a rheumatoid inflammation of the joints”, whilst in the counterfactual context, the extension of the term “arthritis” is “a rheumatoid inflammation of the joints and bones”. Therefore, this difference in extension should theoretically lead to a difference in content and type of the arthritis beliefs of the respective Berts. As Burge puts it:

However we describe the patient’s attitudes in the counterfactual situation, it will not be with a term or phrase extensionally equivalent with ‘arthritis’. So the patient’s counterfactual attitude contents differ from his actual ones (Burge, 1979, p.79).

Therefore, if the “arthritis” attitude contents of actual-Bert and counterfactual-Bert differ, how can there be any sort of content shared between them? However, it is only the objective de dicto reading of counterfactual-Bert’s arthritis belief that we have considered so far. As Loar would put it, Burge has perhaps established that the social content of counterfactual-Bert’s de dicto arthritis belief is different, but there is still



another “layer” of psychological content to consider, what I have called the subjective de dicto reading of the belief. This latter reading of the belief would look like:

Counterfactual-Bert believes(sn) ‘that he has arthritis in his thigh’.

The subjective-notional de dicto reading gives the term “arthritis” the semantic content that Bert seems to give it from his subjective first-person perspective, while of course assuming that counterfactual-Bert does not know the correct meaning of arthritis in his context, just as actual-Bert didn’t in his context. Just like actual-Bert, counterfactual-Bert just believes that “arthritis” is the rheumatoid disease that he has in his ankles and wrists, as well as his thigh, the disease that his father had, etc. The semantic content that counterfactual-Bert’s oblique “arthritis” would have on a subjective de dicto reading is basically something like “a rheumatoid disease in my social environment”. Now, this is pretty much the semantic content that we claimed actual-Bert’s “arthritis” had on a similar subjective de dicto reading, so are we entitled to claim that, contra Burge, it is possible to ascribe beliefs to both Berts using the term “arthritis”? It seems to me that on a subjective de dicto reading one can ascribe arthritis beliefs to the Berts using the term “arthritis” each time.<sup>14</sup> Recall that Burge claimed this was not possible as the extensional concepts of the actual and counterfactual “arthritis” terms were distinct, which meant that the terms should have distinct contents. The upshot of this was that actual-Bert’s arthritis belief was of a different belief-type to counterfactual-Bert’s arthritis belief, so the actual “arthritis” term could not correctly be used in the ascription of an arthritis belief to counterfactual-Bert.

However, there is the possibility that on a subjective de dicto reading of the Berts' arthritis beliefs, their "arthritis" terms, as they themselves use them, do have the same extension conditions for their respective *arthritis* concepts, i.e., the shared concept is roughly "a rheumatoid disease in my social environment". On a de re or objective de dicto reading of the Berts arthritis beliefs, the actual and counterfactual individual concepts of the arthritis diseases are taken into consideration when saying what the semantic content of the "arthritis" terms are, which means that they would be different.

But from the subjective perspectives of both Berts, as they conceive things as being, the semantic content of their respective "arthritis" terms is just the descriptive concept of being the rheumatoid disease in their surrounding social environments. As explained previously, the narrow individuation of their arthritis beliefs means that the extension conditions of their "arthritis" terms are composed of satisfying a certain description, rather than specifically referring to, and including, the individual concepts that make up the respective arthritis diseases. For instance, when narrowly individuated, counterfactual-Bert's "arthritis" term means and refers to only a rheumatoid disease in his social environment; but when widely individuated, counterfactual-Bert's "arthritis" term means and refers to, only the rheumatoid disease of the joints and bones in his social environment (and, *mutatis mutandis*, the same for actual-Bert).<sup>15</sup> However, it must be admitted that the problem of specifying exactly what the narrow conceptual content of the "arthritis" terms is in the above subjective de dicto belief ascriptions is still present.<sup>16 17</sup>

Therefore, in this chapter it has been seen that Loar's ideas about specifying his narrow psychological content are problematic, in that it doesn't seem it can be given a specification using a that-clause. However, I proposed using a so-called subjective de dicto belief ascription to specify the psychological content, particularly in relation to Burge's "arthritis" thought experiment. The next chapter will examine Loar's attempts to provide his psychological content with an intentional aspect.

## Notes to Chapter 2

<sup>1</sup> For example, Edwards claims that "By "commonsense explanations of action" it is meant explanations which ascribe beliefs, desires etc. to subjects which are individuated by that-clauses" (1994, p.24).

<sup>2</sup> For example, at one point Loar declares that "...our view of psychologically explanatory states as ways of conceiving things is not unconnected with our first-person awareness of our own conceptions and the role they apparently play in our behaviour. Just as our own ascription of pain to others involves a *projection* of subjectively apprehended properties, so with psychological content" (1988b, p.135).

<sup>3</sup> Frances (1999) makes the same point, see pp.53-56.

<sup>4</sup> It should be noted that in Burge's (1979) original thought experiment, there is no Bert. Burge refers to a "person" and a "patient"; however, I will stick with using Bert, as he feels part of the family now.

<sup>5</sup> Also as mentioned previously, my use of the ideas of subjective and objective points of view stems from the work of Thomas Nagel (e.g., 1974), which I examine in my 2001 MPhil thesis "Thomas Nagel and Consciousness" which is available from all good chemists.

<sup>6</sup> This is from Sosa (1970), as quoted in Dennett (1982) on p.69.

<sup>7</sup> See Gallois (1996) for a good discussion of many of these different aspects of de dicto and de re beliefs. It is interesting to note here that Sosa's definitions do give the impression that the de dicto and de re distinction is similar to Russell's distinction between knowledge by description and knowledge by acquaintance respectively. Moreover, this impression is strengthened if a subject has to be in the presence of the particular thing (res), in order to entertain a de re belief about that thing. Thanks to my supervisor for pointing this out.

<sup>8</sup> In the example he gives, Burge uses the extensionally equivalent expressions "water" and "H<sub>2</sub>O" to make this point.

<sup>9</sup> As Burge states "The subject need not have entered into special *en rapport* or quasi-indexical relations with objects that the misunderstood term applies to in order for the argument to work. We can appeal to attitudes that would usually be regarded as paradigmatic cases of *de dicto*, non-indexical, *non-de-re*, mental attitudes or events" (1979, p.86).

<sup>10</sup> Bach (1988) also criticises Burge in relation to this point on p.88.

<sup>11</sup> A similar point is also made by Jacob (1990) where he states that:

...even though a word may have an analytic definition, to determine its contribution to the proposition expressed by an utterance of it, one may have to rely on contextual factors such as supplying in thought a mental counterpart of a suppressed expression referring to the particular use of the word in a community or to a person's beliefs (1990, p.160).

Interestingly, Jacob believes that all the beliefs he is considering at that point in his article are de dicto ones (1990, p.156), and in one of the examples he considers, the mental counterpart to the expression "square" is claimed to be "...an abbreviation for some longer utterance involving the words 'what they call "square" in Alfred's community'" (1990, p.160). The latter mental counterpart to "square" is interesting because it has a possible de re ring to it, a point that will shortly come to the fore.

<sup>12</sup> See Block, 1986, p.619 for a clear discussion of the narrow and wide individuation of beliefs.

<sup>13</sup> It can be seen that the reasoning here is similar to Kripke's (1980) notion of the rigid designation of natural kind terms. Kripke took the view that because "water" referred to H<sub>2</sub>O in the actual world, then our actual term "water" would necessarily refer to H<sub>2</sub>O in all counterfactual possible worlds. It was this point that Putnam (1975) used in his famous Twin Earth thought experiment, where he argued that even though on Twin Earth there may be a superficially identical substance to our water, which was also called "water", because it was composed of XYZ molecules rather than H<sub>2</sub>O molecules, if we visited Twin Earth, we could not correctly use our term "water" to refer to it, as we would "mean" something different to what the Twin Earthian term "water" meant. Burge (1979, 1982) himself believes that Putnam did not emphasise the fact that the content of the propositional attitudes containing "water" that could be attributed to Earthians and Twin Earthians would also be different, i.e., the content of their respective water-thoughts would be distinct (perhaps Burge is a little unfair to Putnam here?). The reason for this is similar to the one Burge gives for "arthritis", i.e., one could not correctly use content clauses containing oblique occurrences of the term "water" to attribute propositional attitudes to Twin Earthians.

<sup>14</sup> Perhaps one could query here whether the proposed shared arthritis beliefs of the Berts could properly be called "beliefs". They certainly seem to be mental states of some kind; moreover, it seems to me that both "beliefs" do share some sort of intentionality, as they are both "about" the disease called "arthritis" in the actual and counterfactual contexts.

<sup>15</sup> Is it possible that what I am specifying as the narrow subjective de dicto belief content of actual-Bert and counterfactual-Bert is actually similar to Lewis's (1979, 1994) notion of de se narrow content? For Lewis, de se belief content is not about how the world is, but concerns how the subject himself conceives of things, or rather, it is belief content which is centred on the subject. Moreover, he suggests that de se belief content can be characterised as the subject self-ascribing a property to himself, which will then be true if the subject does indeed possess that property (1994, p.318). The similarity of Lewis's de se content and the subjective de dicto content we have been talking about, comes from the metalinguistic manner which Lewis uses to specify narrow de se belief content. If I understand him correctly, the specification of the narrow de se belief content which actual-Bert and counterfactual-Bert share would go something like this: both Berts would self-ascribe the same properties, i.e., they both self-ascribe the property of having the disease which they have heard is called "arthritis" in their community, in their thighs. To me, this does seem similar to the specifications of the Berts' subjective de dicto "arthritis" beliefs that we have been considering, in that the "arthritis" concepts involved only go so far into the world, they don't go far enough to be individuated by whatever is the specific concept of "arthritis" in the relevant social community.

<sup>16</sup> A possible answer to this problem could be to utilise a suggestion of Crane (1991), which involves drawing a distinction between the conventional meanings of words and the contents of belief (1991, p.19). If I understand Crane correctly, the suggestion is that the actual person in Burge's thought experiment should be ascribed the common concept of tharthritus, which would have arthritis and tharthritus in its extension, as the actual person cannot discriminate between arthritis and tharthritus (1991, p.19). This suggests the following: perhaps we could specify that the shared conceptual content of actual-Bert's and counterfactual-Bert's subjective de dicto "arthritis" beliefs is the disjunctive content of arthritis and tharthritus. This would allow us to give a specification of the narrow content involved, and does seem plausible; as before, actual-Bert and counterfactual-Bert cannot discriminate between arthritis and tharthritus, and, moreover, their behavioural dispositions will be identical in the actual and counterfactual situations. I'm not entirely certain about this, but it might be a possibility.

<sup>17</sup> Some of the issues and conclusions that are contained in the above discussion are similar to ones that have been discussed in some recent literature concerned with Burge's arthritis thought experiment. For instance, perhaps an analogous way to express the above claims, using the terminology of Horowitz (2001), would be to say that counterfactual-Bert's intentional arthritis state is determining the property P, that an object in that world must possess in order to be the intentional object of that intentional state (2001, p.323). In this case, P would be something like a rheumatoid disease in the current social environment. In the case of actual-Bert, the property P that his intentional arthritis state determines, that an object in that world must possess in order to be the intentional object of that state, would be the same as the counterfactual property P. If I understand Horowitz correctly, one might claim that this property P that the intentional state determines in every possible world or context is the same in every possible world or context (2001, p.336), a result which Horowitz takes as proving the falsity of modal externalism (2001, p.336). It does seem that Horowitz's claim is similar to the one that I am making, which is (roughly), that on a subjective de dicto reading of actual-Bert's and counterfactual-Bert's arthritis beliefs, the "object" of those respective beliefs is going to be the rheumatoid disease in their respective social environments. Moreover, this is irrespective of what the individual concept of the

disease “arthritis” is in those environments, i.e., whether “arthritis” is a rheumatoid disease of the joints only, or a rheumatoid disease of the joints and bones. Horowitz’s position is attacked by Brueckner (2003), which then leads to a defence by Horowitz (2005). Will Brueckner attack Horowitz’s defence against his original attack? Only time will tell.

## CHAPTER 3

### LOAR, NARROW PSYCHOLOGICAL CONTENT AND INTENTIONALITY

Having explicated Loar's version of a conceptual role narrow content, viz., his psychological content, it will also be instructive to see how Loar manages to provide his narrow content with an element of world-directedness or intentionality. The problem here is that if it is the wide content of a thought that has referential truth conditional content, then what does the actual "content" of narrow content consist of? It is all very well to say that psychological explanation or conceptual role may *individuate* beliefs differently from their oblique that-clause ascriptions; but what is the *content* of these differently individuated beliefs? (Loar, 1988a, p.105). Loar sets out the commonly held objection to narrow content, namely:

...that so-called narrow content cannot capture an intentional property; for the two beliefs in the Burge case and those in the Twin Earth case do not share truth conditions. "Content" should mean intentionality, and intentionality is a certain directedness of thoughts onto things, properties, states of affairs, in short, truth conditions and the components of truth conditions (Loar, 1988a, p.107).

As Loar says above, the two beliefs in the Twin Earth example have different truth conditions. However, both these thoughts would be considered as *intentional*, as they are directed at certain natural kinds or substances in the respective contexts, i.e., H<sub>2</sub>O and XYZ. This also means that if a necessary condition of being a contentful state is

having intentionality, then both these thoughts can be said to have content. However, this is a problem for narrow content, for it is supposed to be a content that me and my doppelganger's beliefs have in common. But the common "content" cannot include the extensions of our "water" beliefs, because these are considered different.

Therefore, it will be difficult to consider narrow content as intentional, as it is the extensions of the beliefs which are the "directed onto the world" parts of the beliefs.

However, is this an accurate appraisal of the situation? Even though me and Twin-Me's respective water-beliefs might be said to have different wide contents, and so different truth conditions, it could still be claimed that the shared narrow content of our respective water-beliefs has an intentional element to it. From the respective subjective viewpoints of Twin-Me and I, we would take ourselves to be referring to, and have a concept of, the clear, colourless, tasteless, liquid, which we call "water" in our respective contexts, and which has a certain conceptual role in our lives. Now, it seems plausible to say that this narrow content would also be intentional, in that it is world-directed or about something in our respective environments. So, despite the different truth conditions of our water-beliefs, the narrow content of me and Twin-Me's water-beliefs could still seem to possess an element of intentionality. Indeed, perhaps one could wonder here that with the above mentioned objection, whether "intentionality" and "truth conditions" are equivalent? That is, perhaps mental states could still have a shared intentionality despite their having different truth conditions. This point is a pertinent one as we now examine Loar's suggestions for answering the "intentionality" objection to narrow content.

### 3.1 Loar's Journey to his "Context-Indeterminate Realization Conditions"

Loar's first suggestion is that narrow content could be a content that does not involve truth-conditions. He points out that in the earlier Bert, Pierre and Paul examples, when thoughts are individuated by their conceptual roles, this individuation is seen to be quite different and separable from the truth-conditions of the thoughts (Loar, 1988a, p.107). For instance, two beliefs which have different truth-conditions (as these were judged from their oblique that-clauses), are considered the same, as individuated by their conceptual roles. Now, if truth-conditions are not going to apply to the psychological content of the beliefs as individuated by conceptual role, then the view could be taken that this content is not going to be intentional, in the sense of being "about" something in the world. This is going to be problematic if one takes the view that a belief can only have content if it has intentionality.

However, as Loar suggests, we do not have to accept this as a necessary condition for a belief to have a content. Even though the narrow psychological content of a belief may not necessarily have truth-conditions, it still seems to have a "content" to the person whose belief it is.<sup>1</sup> That is, from the subjective point of view of the person concerned, their narrow belief does have a content, which comprises how they perceive and conceive the world to be (Loar, 1988a, p.108).<sup>2</sup> For example, from the subjective viewpoint of Bert, he may have a belief with the content "I have arthritis", for that is how he conceives things to be. This is despite the fact that he may be in the counterfactual situation where it may be more appropriate to ascribe to him the belief that he has tharthritis. Even though factors in Bert's social environment have changed, the psychological content of his belief has remained the same, and it does



appear to be some sort of content. Furthermore, despite this content not involving truth-conditions, it does seem that intentionality is involved, as Loar points out:

...the conceptual roles of one's own thoughts determine how one conceives things, and it is difficult to see how one can consider how oneself conceives things without that in some sense involving what one's thoughts are "about" (Loar, 1988a, p.108).

The psychological content of a belief may well have been *caused* by interactions with the world, but it need not necessarily match up with how the world actually is at a particular time. Again, subjective and objective perspectives in belief ascription may be appropriate here. The subjective psychological content of a belief may not be identical with the objective social content of the belief, as that is ascribed in an oblique that-clause. It is also the oblique that-clause description of the belief that is taken to give the truth-conditions of that belief; but as this may only contain social content, then it could be claimed that these are the "socially determined truth-conditions" of the belief in question (Loar, 1988a, p.110). From the objective third-person perspective there appears to be no subjective content to a belief, only objective content; whereas, from the subjective first-person perspective, the belief may well have a discernible content that is distinct from its "official" objective social content.<sup>3</sup>

If Loar is adopting the above point of view, it seems that there cannot be an appearance/reality distinction with regard to the narrow mental content of a subject's thoughts. For example, in the counterfactual situation of Burge's "arthritis" example, the narrow content of Bert's belief that he has arthritis in his thigh, is simply that he has arthritis in his thigh. The way Bert conceives the situation, from his subjective viewpoint, is that he has arthritis; this is despite the fact that the social content of his

belief, as specified in a that-clause, might lead to Bert being ascribed the belief that he has tharthritis in his thigh. A similar case could be made for the Twin Earth scenario involving H<sub>2</sub>O and XYZ. If I left the Earth and landed on Twin Earth, the narrow content of my water-beliefs as individuated by conceptual role would be exactly the same as they were on the Earth. From my subjective conception of the situation, I would still be drinking and swimming in, etc., the clear, odourless, etc., liquid called water, which I had previously done back on Earth. This is despite the fact that the wide content of my water-beliefs would be different on Twin Earth (XYZ instead of H<sub>2</sub>O), which would lead to me being objectively ascribed Twin water-beliefs instead of water-beliefs. Of course, I think that this is a more arguable claim than in the arthritis example, what with Kripke's (1980) and Putnam's (1975) persuasive views on natural kind concepts, but it still seems fairly plausible to me. From the subject's first-person perspective, the *apparent* content of his beliefs *really is* the content of his beliefs; despite the fact that the extension of those beliefs, and so the truth conditions of those beliefs, entail that the subject has beliefs about the world which are false.

### 3.1.1 *The Realization Conditions of a Belief*

Loar's next suggestion is that if narrow content must have some intentional aspects, then *realization conditions* could be used instead of truth-conditions. As previously explained, the usual way of explicating the intentional aspect of a belief-content is through its truth-conditions, i.e., if the belief has a that-clause which correctly ascribes it, then it is intentional, in the sense that it has a content which is "about" an actual state of affairs in the world (Loar, 1988a, p.108). Now, narrow content, which is

viewed as content which is shared between doppelgangers, or the same person in a counterfactual situation, cannot acquire its intentionality in this way, as in Putnamian and Burgean thought experiments, the relevant worldly states of affairs are usually different, which thus means that the truth-conditions of the belief-contents are also different.

However, Loar comes up with another method to give narrow content a form of intentionality, which is to use the *realization conditions* of the narrow belief-content. What Loar means by this is best explained by using an example. Imagine that Bert (he of rheumatic fame), writes a note in his diary along the following lines, “My ankles are hurting, I think I must have arthritis, so I will catch the bus into town and see the emergency doctor”. Now, as Loar points out, if we are just reading this entry from Bert’s diary, not knowing whether he is in the actual or counterfactual situation, i.e., not knowing the facts about Bert’s social environment when he uses the word “arthritis”, then we do not know if his use of the word is a misconception or not (Loar, 1988a, p.108). However, Loar contends that we do know:

...how the world would be if Bert’s conceptions are or were not misconceptions. How Bert thinks of things – as that is described from an individualist perspective – appears to determine a set of possible worlds; namely, those in which Bert’s thoughts are or would be true if they are or were not misconceptions. Call that set of worlds the *realization conditions* of Bert’s beliefs (Loar, 1988a, p.108).

So, it can be seen here that Loar is trying to give Bert’s belief-content (which is a narrow content, because Bert could be in the actual or counterfactual situation) an intentional aspect through the use of the realization conditions of Bert’s belief.

However, I am not sure about this strategy. Does Loar mean that there is a possible world in which Bert has the true belief that he has arthritis, and this would be intentional in the sense that there would be an actual worldly state of affairs, i.e., Bert's having arthritis, which his belief was directed towards? This is certainly an interesting move by Loar. Whereas, normally it is truth-conditions which beliefs are directly connected with (whether they are true or false), he gets to a truth-condition for Bert's belief indirectly, via the realization conditions.<sup>4</sup> He treats Bert's belief more like it was a desire, which is normally associated with satisfaction conditions rather than truth conditions. For example, if Bert desires that his arthritis be cured, then his desire also seems to determine a set of possible worlds in which his desire is satisfied, rather than unsatisfied; and this set of worlds could be called the satisfaction conditions of Bert's desire. And again it could be claimed that Bert's desire has an intentional aspect to it, as, if satisfied, it would be directed onto an actual state of affairs in the world, i.e., Bert with his arthritis cured.

### 3.1.2 *The Context-Indeterminate Truth Conditions of a Belief*

However, the use of realization conditions is not quite Loar's final resting place in his considerations on elucidating the intentionality of narrow content. Loar eventually takes the view that if truth-conditions cannot be used, then context-indeterminate realization conditions may be just right for the job. To get to this stage, he first considers the notion of context-indeterminate truth-conditions, which he introduces with the following situation:

Suppose I find a diary with the entry: “Hot and sunny today; phoned Maria to invite her to the beach”. Now, the date has been torn off the page. Still I appear to understand the diarist’s explanation of his/her phoning Maria, despite not knowing the truth conditions (in one sense) of the thought expressed by “hot and sunny today” (Loar, 1988a, pp.108-109).

Now, Loar says that there is “one sense” in which we don’t know the truth conditions of the thought “hot and sunny today”, but of course there is one sense in which we do know the truth conditions. Presumably, the truth conditions of the thought expressed by “hot and sunny today” depend on whether the day the diary entry was made, it was actually hot and sunny: if it was, then the thought was true, while if it wasn’t, then the thought was false. I suppose the “sense” in which we don’t know the truth conditions is that we don’t know the exact day that the diary entry was made, the truth conditions are not anchored to a particular context, they are “floating free”, so to speak.

However, as Loar declares, this lack of a specific context for the truth conditions does not obstruct our commonsense psychological explanation of why the diarist behaved as he/she did (just as the lack of a specific context for the truth conditions did not obstruct our psychological explanation of why Bert caught the bus into town, in the previous suggestion). Furthermore, even though the thought that “it is hot and sunny today” when thought on two different days would usually be judged to have different truth conditions, both thoughts might be said to have the same truth conditions, if one abstracts away from the particular day or context that the thought was thought, i.e., the shared truth conditions would be “it is hot and sunny today” (Loar, 1988a, p.109). Loar calls an abstracted truth condition of this sort *context-indeterminate*, by which he means that it is:

...a function that maps the contexts [that two thinkers and their beliefs] respectively inhabit onto the sets of possible worlds that constitute the “context-determinate” truth-conditions of their beliefs (Loar, 1987, p.95).

As we shall see, this abstraction of truth conditions away from context is similar to Fodor’s (1987) abstraction of thought-content away from the two different contexts in which two thoughts occur, in order to produce the narrow content that both thoughts share. For example, if I think that “water is wet” and Twin-Me thinks that “water is wet”, and if one is persuaded by the Twin Earth intuitions, then these thoughts would be considered to have different truth conditions, as I am referring to H<sub>2</sub>O and Twin-Me is referring to XYZ. However, Fodor claims that it is possible to abstract away the narrow thought-content that “water is wet” from the contexts in which the thoughts actually occurred. Perhaps Loar’s context-indeterminate truth conditions might also be called “narrow” truth conditions, which have been abstracted from the “wide” truth conditions, which would be context-determinate? Indeed, if I have the thought “water is wet” on Earth and Twin-Me has the thought “water is wet” on Twin Earth, wouldn’t the abstracted context-indeterminate truth conditions be something like “water is wet”? I am not sure if this is what Loar means, but in his 1988 article, he certainly believes that if the narrow content utilised by psychological explanation is in need of a truth conditional aspect, then context-indeterminate truth conditions are a possible contender (Loar, 1988a, p.109).

### 3.1.3 *The Context-Indeterminate Realization Conditions of a Belief*

Nevertheless, Loar’s preferred idea for giving an intentional aspect to narrow content is that of context-indeterminate realization conditions, which can be viewed as a

synthesis of the two previous ideas looked at. Loar does not exactly elaborate on these conditions, but he does give some clues as to what he means in a footnote, where he briefly discusses Fodor's (1987) views on narrow content. Fodor held the view that narrow content could be construed as a function from contexts and thoughts onto truth conditions or possible worlds. Loar criticises Fodor's notion, where he claims that Fodor should not "...treat natural kind terms as if they are pure indexicals whose semantic values are determined by context..." (Loar, 1988a, p.n197), as this results in "water" and "alcohol" having the same function, and so, the same narrow content. In contrast, Loar believes that his context-indeterminate realization conditions would narrowly individuate beliefs about "water" and "alcohol" as distinct, according to their different conceptual roles, as his aforementioned conditions are supposed to include a "...combination of indexicality and substantive conceptual content" (Loar, 1988a, p.n197).

So, perhaps what Loar means is the following. Consider Bert again, and his belief that he has arthritis in his ankles. As stated previously, we do not know what sort of social environment Bert is situated in, so we cannot say whether his belief is a misconception or not. But what we can do is abstract the realization condition "Bert has arthritis in his ankles" from the context in which it took place, in a similar manner to the way the truth condition "hot and sunny today" could be abstracted from its context in Loar's diary example. So, the realization condition is then context-indeterminate, it is not anchored to any particular context. However, it is important to note that the realization condition is still anchored to the particular belief whose realization condition it is, i.e., Bert's actual belief; also, the realization condition will be anchored to any conceptual connotations that the belief had for Bert. So, the

intentionality that this belief can gain is achieved by imagining a possible situation or world in which the realization condition of this belief is satisfied, which in turn entails that the condition becomes anchored to the particular context of that possible world. And of course, the possible world in which the realization condition is satisfied will be one where Bert's belief that he has arthritis in his ankles is not a misconception, and so is true.

In saying this, I don't think that Loar's context-indeterminate realization conditions are affected by Bert's causal history. The situation being imagined is that we don't know whether Bert is on the Earth or Twin Earth, or in the actual or counterfactual situation. That is, we don't know whether he is Earth-Bert or Twin Earth-Bert, or actual-Bert or counterfactual-Bert. But whoever he is, Bert has grown up in that context, we are not considering a situation where Earth-Bert has been transferred to Twin Earth, for example (at least, I don't think we are). Therefore, when the realization conditions of his arthritis-belief have been satisfied, it has to be a context where it is true that Bert has arthritis in his ankles, and his causal history is not a problem, as he is in the social context that he has always been in. I am not sure if this is what Loar has in mind, but it does seem plausible to me.



## 3.2 Assessment and Criticisms of Loar's Realization Conditions

### 3.2.1 *An Explication of Realization Conditions in Terms of Possible World*

#### *Semantics*

A critical analysis of Loar's thesis of realization conditions is done by Stalnaker (1990), where he frames the issues in terms of possible world semantics:

Imagine a set of possible worlds, in all of which Bert exists and is thinking a certain thought. Now we can distinguish the following two questions: first, we may ask what the actual content of Bert's thought is, and then evaluate that content relative to each of the possible worlds in the set. Second, we may ask, for each of the worlds in the set, what the content of Bert's thought is in that world, and whether the content is true in that world (Stalnaker, 1990, p.197).

As Stalnaker goes on to say, this means that with the first question, the content of Bert's thought is fixed as whatever it is in the actual world, then this fixed-content thought is evaluated at each of the worlds in the set for its truth-value. In contrast, with the second question, the content of Bert's thought is not fixed by the actual world, but the content that the thought has in each of the possible worlds is evaluated for its truth-value in that world (Stalnaker, 1990, p.197). To explain how all this is supposed to specify a narrow content, Stalnaker then gives an example involving "water". I will follow the structure of his example, but will use the term "arthritis" instead of "water".

Imagine that Bert has the belief "I have arthritis in my ankles". Following Stalnaker, we can evaluate this belief relative to four possible worlds: *a* is the actual world, where "arthritis" is defined as a rheumatoid ailment of the joints (we can also suppose that Bert really does have arthritis in his ankles); *b* is the counterfactual Twin-Earth

world, where “arthritis” is defined as a rheumatoid ailment of the joints and the bones (perhaps we could call it “tharthritis”), once again we can suppose that Bert really does have arthritis in his ankles; *c* is a world just like the actual world (where arthritis is a rheumatoid ailment of the joints only), but where Bert does not really have arthritis in his ankles, he had actually sprained them while doing some over-vigorous jogging on the spot to impress his next door neighbour; *d* is the same scenario as *c*, except that arthritis is defined as a rheumatoid ailment of the joints and bones.

So, in each of these four worlds Bert has the belief “I have arthritis in my ankles”: how do the respective evaluations of his belief turn out? In world *a*, which is the actual world, Bert’s belief is true, as he *actually* does have arthritis in his ankles. Following Stalnaker, we can say that the true proposition that comes from Bert’s belief in the actual world is the “actual wide content” of his thought (Stalnaker, 1990, p.198). If we follow the method of the first question which was mentioned above, and fix the content of Bert’s belief to what it is in the actual world, and then evaluate the proposition of that fixed-content belief in each of the other worlds, then we shall find that it is false in *b*, *c* and *d*. It is straightforwardly false in worlds *c* and *d*, because in those worlds Bert does not have arthritis in his ankles, he has misconceived his situation in each of those worlds. However, in world *c*, even though his belief is false, the wide content of the belief is the same as it is in the actual world, as in *c* “arthritis” is viewed as being a rheumatoid ailment of the joints only (Stalnaker, 1990, p.198). When Bert’s fixed-content belief is evaluated at world *b*, it will also be false. The reason for this, if we go along with the Burgean intuitions, is that for Bert’s belief to have a true proposition in *b*, it would have to take the form “I have tharthritis in my ankles”. This difference in propositions of the respective beliefs in *a* and *b* is

supposedly due to the differences in the social and linguistic environments of the respective worlds (i.e., the different conceptions of arthritis in each world).

Therefore, when we ask the first question, fixing the content of Bert's belief in the actual world, and then evaluating this fixed-content belief in relation to the other possible worlds, there does not seem to be any conception of narrow content that emerges.

However, what happens when we ask the second question? That is, for each of the possible worlds, we ask what the content of Bert's belief is in that world, and what the truth-value of the belief-content would be when evaluated at that world. So, let us again imagine that in each possible world Bert has the belief that "I have arthritis in my ankles". It seems fairly clear that, as before, Bert's belief will be false in worlds *c* and *d*, as in these worlds he does not have arthritis in his ankles, but has misconceived the situation in each world. But what about worlds *a* and *b*? Well, the content of Bert's belief in world *a* (the actual world) is that he has arthritis in his ankles, and it seems that when this content is evaluated at *a*, it will be true, as in that world Bert does have arthritis in his ankles. Therefore, in world *a*, Bert's belief will produce a true proposition. The content of Bert's belief in world *b* (the counterfactual world) is also that he has arthritis in his ankles, and it also seems that when this content is evaluated at *b*, it will be true, as in that world Bert has arthritis in his ankles.

Therefore, in world *b*, Bert's belief will produce a true proposition. So it can be seen that the second question seems to produce the same truth-values for Bert's belief in worlds *a* and *b* (the actual and counterfactual worlds), i.e., they are both true. This is contra to the conclusion of the first question, which resulted in Bert's belief being true in *a* and false in *b*. Why the difference?

The difference of course depends on the fact that with the first question the content of Bert's belief that he has arthritis in his ankles is fixed as whatever it is in the actual world. If we go along with externalist intuitions, this has the effect of making the content of Bert's belief "I have arthritis in my ankles" something like "I have a rheumatoid ailment of the joints in my ankles". In fixing the content of Bert's arthritis belief to what it is in the actual world (*a*), this entails fixing the concept of "arthritis" to whatever it is in the actual world, i.e., a rheumatoid ailment of the joints. This is analogous to the point that Kripke (1980) and Putnam (1975) made with regards to natural kinds such as water. Once we find out that "water" is H<sub>2</sub>O in the actual world, then the term "water" becomes a rigid designator, and denotes H<sub>2</sub>O in all possible worlds. This means that on Twin Earth we cannot call the clear liquid in the oceans and rivers "water", as it is composed of XYZ and not H<sub>2</sub>O. It is not water as we know it, but only "watery stuff", which we can call "twin-water", "twater", etc.

The same is supposedly true of the rigidified term "arthritis" in Bert's belief. When Bert's fixed-content belief "I have arthritis in my ankles" is evaluated at the counterfactual world (*b*), its truth-value is judged to be false. The reason for this is that on *b* the term "arthritis" has a different concept than it does on the actual world, i.e., on *b* the concept of "arthritis" is that it is a rheumatoid ailment of the joints and bones. For Bert's belief to be evaluated as true in relation to world *b* it would have to be something like "I have tharthritis [a rheumatoid ailment of the joints and bones] in my ankles". But of course, this not the content that Bert's belief has, its rigidified term "arthritis" makes the content of the thought come out as "I have arthritis [a rheumatoid ailment of the joints] in my ankles", which, in relation to world *b*, would



be evaluated as false (so the externalist intuitions tell us). It is as if when a term in the actual world becomes a rigid designator of something, then whenever that term is evaluated in a different possible world, it has dragged a social or physical piece of the actual world with it, which determines its meaning, irrespective of the environment of the world it is now being evaluated at.

However, with the second question, the content of Bert's belief that he has arthritis in his ankles is not fixed by the actual world, but rather the content that the thought has in each world is evaluated for a truth-value in relation to that world. Put another way, it could be said that with this procedure we don't know which world is the actual one that Bert's belief is going to be evaluated in. Even though *a* is called the actual world, Bert's belief could well be evaluated at world *b*, which is the counterfactual world. If this did occur, then world *b* would be considered as Bert's actual world, with interesting consequences. For example, if *b* was the classic Twin Earth world where water was composed of XYZ, then as far as Bert would be concerned, Twin Earth would be his actual world and "water" would really be XYZ, and not H<sub>2</sub>O.

The effect of this procedure on Bert's arthritis thoughts in worlds *a* and *b* would be the following. If Bert's thought that he has arthritis in his ankles is evaluated at world *a*, it will be true. The reason for this is that on *a* the concept of "arthritis" is that it is a rheumatoid ailment of the joints. This means that the content of Bert's belief that he has arthritis in his ankles, will be that he has a rheumatoid ailment of the joints in his ankles. Therefore, when this belief is evaluated in relation to world *a*, it will be a true belief. Furthermore, when Bert's belief that he has arthritis in his ankles is evaluated at world *b*, it will also turn out to be true. The reason for this is that on *b* the concept

of “arthritis” is that it is a rheumatoid ailment of the joints and bones. This means that the content of Bert’s belief that he has arthritis in his ankles, will be that he has a rheumatoid ailment of the joints and bones, in his ankles. Therefore, when this belief is evaluated in relation to world *b*, it will also be a true belief. Therefore, despite Bert’s arthritis beliefs in *a* and *b* having different wide contents (due to the different social concepts of “arthritis” in *a* and *b*), there does seem to be some sort of narrow content involved in both beliefs when we follow the procedure of the second question.

Indeed, I think that Stalnaker’s conception of narrow content that he finds using this particular possible worlds procedure, is quite analogous to Loar’s idea of realization conditions for narrow beliefs and other attitudes. It will be recalled that Loar introduced his idea of realization conditions to try to indirectly bring some intentionality to the narrow content shared between physical duplicates. With the wide content of beliefs this is normally done with truth conditions, but of course in the Putnamian and Burgean thought-experiments, the beliefs of the physical duplicates always seem to have different truth-values, because of differences in the physical or social environments. However, let us now go back to Loar’s original example of Bert’s entry in his diary, which says that he believes he has arthritis in his ankles and is going to see the doctor. By just reading his diary, we don’t know whether Bert is in the actual world or the counterfactual world, i.e., we don’t know whether the public concept of the term “arthritis” is a rheumatoid ailment of only the joints, or of the joints and bones in general. This is equivalent to the above situation where we don’t know whether it is world *a* or *b* that is the actual world where the content of Bert’s belief is going to be evaluated.

However, we do know the following: if Bert's belief is not a misconception (i.e., he has not mistaken sprained ankles for arthritic ankles), then the world in which Bert's belief would be true, is one in which he would have arthritis in his ankles. And this would be true whether Bert's condition is realized in the actual or counterfactual world, which means that the content of Bert's belief would be determined in relation to the world he is located in, and then evaluated in relation to that world, i.e., there would be no rigidified content involved with Bert's arthritis belief. This would have the result that Bert's belief would have the same truth-value in the actual and counterfactual worlds, i.e., it would be true in both cases. This is equivalent to the above situation as envisaged by Stalnaker, where the content of Bert's arthritis belief, when determined and evaluated relative to the possible world it is in (either *a* or *b*), produces a true belief in each case. Therefore, Loar's thesis of realization conditions seems to produce the same truth-values for narrow belief contents; which, in turn, seems to endow those narrow belief contents with some intentionality.<sup>5</sup>

### 3.2.2 *Stalnaker's Critique of Loar's Realization Conditions*

In addition to setting out Loar's thesis of realization conditions in terms of possible worlds, Stalnaker also has some more general criticisms for Loar's position. Firstly, Stalnaker doubts whether it is right to presuppose that a belief or other attitude can be individuated or identified independently of its wide content (Stalnaker, 1990, p.198). It is this presupposition that he says Loar's realization conditions involve, in that they are context-indeterminate, i.e., abstracted away from a context in which, say, the belief that Bert has arthritis in his ankles, does have a content. He claims that doing

this with a belief expressed in a speech act is unproblematic, but that it is not so obvious to gauge what content an abstracted belief would have had on Twin Earth (Stalnaker, 1990, p.198).

Loar's realization conditions certainly do depend on abstracted beliefs and other attitudes; when Bert believes that he has arthritis in his ankles, we don't know whether he is on Earth (where "arthritis" is a rheumatoid ailment of the joints only) or on Twin Earth (where "arthritis" is a rheumatoid ailment of the joints and bones).

Now, according to Stalnaker, if Bert *utters* the sentence "I have arthritis in my ankles", then there is no problem in abstracting this sentence from whatever context it is uttered in, and evaluating the content that it has in other contexts. Presumably, this works because the word "arthritis" in Bert's utterance is being used with *semantic deference*, i.e., Bert is intending to use the word "arthritis" with the same meaning or concept that others attribute to it in the social context in which Bert is located.

Another way to put this is that Bert intends his use of the word "arthritis" to refer to the "same phenomenon" that the social community's use of the word "arthritis" refers to (Chalmers, 2003a, p.62).<sup>6</sup> This means that the extension of the word "arthritis" as used by Bert, will be the same as the extension of the word "arthritis" as used by whichever social community he is embedded in, i.e., Bert and the social community will have the same extension-conditions for the term "arthritis". So, if Bert is embedded in Earth's social community, the extension of his word "arthritis" will be "a rheumatoid ailment of the joints only"; while, if he is embedded in Twin Earth's social community, the extension of his word "arthritis" will be "a rheumatoid ailment of the joints and bones".



Now, the reason I am labouring over this, is that if it seems fairly straightforward for the utterance of a belief to be individuated independently from its content in the above manner, surely the same can be said of the belief behind the utterance. After all, Bert is meant to be *expressing* his belief that he has arthritis in his ankles with his utterance “I have arthritis in my ankles”. Stalnaker admits that:

If we assume that there is a language of thought, and that to have a belief or desire is to have a token of a mental sentence expressing the belief or desire stored in the appropriate place in the brain, then the presupposition would be unproblematic...(Stalnaker, 1990, pp.198-199).

However, I don't think one needs to assume a language of thought in the manner Stalnaker suggests. When Bert utters the sentence “I have arthritis in my ankles”, I don't think that it would be problematic to further assume that Bert also has a corresponding token belief-state with the content “I have arthritis in my ankles”. Furthermore, it seems plausible to me that Bert's belief-state could also be abstracted from the context it occurred in, and its content evaluated in relation to other contexts.

Now, to this, Stalnaker could claim “...we have no obvious way to individuate the property of believing that *P* except by using the content, that *P*” (Stalnaker, 1990, p.199). However, it seems to me here that Stalnaker is presupposing belief individuation by wide content. Couldn't an equally legitimate internalist presupposition be made here that Bert's arthritis-belief consisted of an internal state, such as an intension, which would determine an extension in a given context? Of course, it could be acknowledged that the creation of Bert's intension is due to external factors, but once the arthritis-intension is formed in Bert's mind, the content of that intension could be evaluated in different contexts. Therefore, if the intension

is something like “I have the disease called arthritis in my ankles”; whether Bert is on Earth or Twin Earth, he would still intend his use of the concept “arthritis” in his arthritis-belief to semantically defer to the concept of “arthritis” which is conventionally used by whichever social community he is embedded in.

There is another alternative, one which is suggested by Stalnaker himself, where he says that it might be possible to “...identify occurrent thoughts independently of their contents” (Stalnaker, 1990, p.198). An “occurrent thought” is usually defined as a thought which is consciously before the mind of a subject at a particular time (in contrast to, say, forgotten thoughts, subconscious thoughts, etc.). Bert’s belief in Loar’s realization conditions example, might be viewed as an occurrent thought of Bert’s. Imagine that Bert is writing the entry “I believe that I have arthritis in my ankles...” in his diary. Now, at a particular moment, he stops writing, looks out of the window, and has the occurrent thought that he does have arthritis in his ankles. At this point, we can ask what the realization conditions of Bert’s occurrent thought *are*. The upshot of this is that initially, Bert’s occurrent arthritis-belief doesn’t have a content, as we are not sure which world is going to be the actual world. It is only when Bert’s belief is realized, that is, when the belief is embedded into a world where it will be true, that it then gains a content.

Another objection that Stalnaker raises involves the question of whether Loar’s notion of realization conditions could deliver a “determinate narrow content”, i.e., if they could determine by a reliable procedure, what exactly the narrow content shared between two beliefs actually is (Stalnaker, 1990, p.199). He believes that Loar’s notion of narrow content “...provides no basis for saying when narrow content

differs, or changes, and when it remains the same” (Stalnaker, 1990, p.199). If belief tokens are identified by their syntactical or physical properties, according to Stalnaker, this would not be sufficient to determine “...the narrow content of the thought token”; since we could imagine a situation where the same physical property that determined the original thought, had a different wide and narrow content, due to differences in the “functional organization of the thinker” (Stalnaker, 1990, p.199).

Stalnaker then gives an analogy of this situation in relation to sentences:

Suppose that in a counterfactual situation more radically different from ours than Twin Earth we find the pattern of sounds identical to the one that in fact constitutes an utterance of [“I have arthritis in my ankles” (this is the example I used)]. But suppose that in this counterfactual world, these sounds occur in a language utterly unlike English, both in syntax and semantics. What they mean there might be roughly translated as “what time does the next bus leave for the zoo?” Presumably, the similarity in the acoustical pattern will not be sufficient to give this utterance the same meaning or content that it has in the actual world in any sense, wide or narrow, even though one could define a propositional concept, and a diagonal proposition representing realization conditions, for a sentence token individuated simply by its sound pattern or shape (Stalnaker, 1990, p.199).

This is an interesting objection from Stalnaker, though I’m not sure how effective it is against Loar’s realization conditions. It seems that what Stalnaker is saying is that even if one could individuate a belief independently from its context, the narrow content that resulted may not be of a determinate kind for two reasons: Firstly, due to changes in the “functional organization of the thinker”, the same particular physical state that is said to cause the belief token “I have arthritis in my ankles”, may have a completely different wide and narrow content in another situation. Secondly, the narrowly individuated belief “I have arthritis in my ankles” (which, at this point is minus any fixed or anchored wide content) could result in realization conditions for the belief at worlds where the belief “I have arthritis in my ankles” means something

completely different to what it does in the actual world. So if the belief is individuated by word-forms as above, this might not be sufficient to ensure that the belief has the same narrow and wide content at various possible worlds where it is realized.

Stalnaker's first reason to doubt a determinate narrow content is, I think, not too serious an objection to realization conditions. For whether we use a doppelganger of Bert or imagine Bert in a counterfactual world, all the Berts involved are going to be physically identical. A doppelganger of Bert is an internal physical duplicate of Bert, which usually entails that as well as having a neurophysiological identity shared between them, Bert and his doppelganger also share a functional identity between them. When Bert himself is considered as existing in a counterfactual world, this "twin" of Bert is also considered to be neurophysiologically and functionally identical to the original Bert, as it seems necessary that Bert is internally physically identical to himself. If this is accepted, then it seems to follow that the narrow content of a particular belief of Bert's will have a determinate content, or at least, a content that is not affected by a different functional organization of beliefs in a twin-Bert.

Indeed, Stalnaker himself seems to accept that the mental states of "internally indiscernible twins" will have the same narrow contents (Stalnaker, 1990, p.199), but then goes on to say that subjects who were thinking the same thoughts, but who were not internally indiscernible, would also be expected to have narrow thought-content shared between them, and that it would be problematic to work out exactly what the content is that they shared. Now, Stalnaker might be quite right about this latter point; however, perhaps we should take stock of the situation we are now faced with.

Stalnaker has basically acknowledged that with the use of realization conditions, the narrow belief-content shared between internal physical duplicates would be the same, which seems quite a result for desperate internalists everywhere. Any further problems might be affected by one's views of narrow and wide content in general. For example, we don't have to agree with Loar that that-clauses generally contain wide or social content. Perhaps we could argue that content in general is narrow, but that in Putnamian or Burgean situations, there would be a wide aspect to thoughts as well as the narrow aspect that is normally there. In these situations we could use something like realization conditions to show what the narrow aspect of the thought consisted of, e.g., perhaps using logical symbols or something similar.

Stalnaker's second reason to doubt a determinate narrow content is a tricky one. It does seem plausible that Bert's belief that he has arthritis in his ankles could have realization conditions in a world where his narrow belief-content would mean something completely different; for example, perhaps the concept of "arthritis" on that world would have an extension something like "the class of blue furry toys". So, when Bert's belief "I have arthritis in my ankles" is realized on that particular world, his use of "arthritis" would semantically defer to the concept of "arthritis" that is used in that social environment. The result of this would be that on that possible world, Bert's belief would be that he had blue furry toys in his ankles. Whilst this belief might also necessitate catching the bus into town to seek medical attention, it would have a different narrow and wide content to that of Bert's original belief. What seems to be needed here is for the realization conditions of a belief to have some element which constrains the worlds where the belief can be realized. However, does Loar's

conception of realization conditions already have an element like this contained in them?

Let's go back to Bert and his original belief that he has arthritis in his ankles.

According to Loar, we can imagine a set of possible worlds where, if the belief is not a misconception, Bert's belief would be true, i.e., Bert would have arthritis in his ankles. This means that if Bert's belief is evaluated in two worlds where the concept of "arthritis" at each has different extensions (e.g., on Earth "arthritis" would be a rheumatoid ailment of the joints only, while on Twin Earth "arthritis" would be a rheumatoid ailment of the joints and bones), as long as Bert really did have arthritis in his ankles in either world, then his belief would be true in both worlds. The reason for this would be that Bert would intend his use of the word, and concept of, "arthritis", to semantically defer to the way the word and concept of "arthritis" is conventionally defined and used in the social environment in which his belief is realized. But of course, this would mean that if Bert's belief is realized on a world where the conventional concept of "arthritis" is somewhat bizarre, e.g., where it meant blue furry toys, then Bert would still intend his use of "arthritis" to semantically defer to the surrounding social environment, with the result that he would have a belief that he has blue furry toys in his ankles.

However, is this really plausible? The realization conditions of Bert's arthritis-belief would result in a *true* belief at a possible world. I don't think that we could plausibly say that Bert's arthritis-belief at the "blue furry toys" possible world would be a true one. So, there does seem to be some constraining element already built into the notion of realization conditions, i.e., it would have to be a true belief that is realized.

This in turn seems to mean that when Bert is having his original arthritis-belief, there must be experiential and conceptual role elements involved in his having the belief, so that he has a pretty good idea of what is involved when he believes that he has arthritis in his ankles.<sup>7</sup> Put another way, Bert's subjective conception of his situation must play a part in identifying his arthritis-belief. Roughly, Bert is experiencing a pain in his ankles, which he believes is caused by the rheumatoid ailment called "arthritis" in the social environment in which he is located. Therefore, Bert's subjective conception seems to exclude the possibility of his belief being realized in the "blue furry toys" world. So it does seem plausible to suppose that Loar's notion of realization conditions does have built-in subjective experiential and conceptual role elements, which constrains the worlds that a belief can be realized in.

If this supposition is not accepted, then another alternative would be to include an extra experiential element to the realization conditions of a belief to achieve the desired effect. A prime candidate for this job could be Edwards' (1994) idea of an "experiential component" or "E-component". Edwards uses his E-component, not in relation to narrow content, but in relation to the individuation of demonstrative thoughts (Edwards, 1994, p.161).<sup>8</sup> Demonstrative thoughts are referring thoughts that usually take the form of "*this* F" or "*that* G", so that they have an indexical or "context-sensitive" element contained in them (Edwards, 1994, p.159; Yourgrau, 1990, p.1). The E-component is used to supplement the use of that-clauses in describing demonstrative thoughts, as that-clauses are shown to be inadequate to completely specify demonstrative thought-content (Edwards, 1994, p.162). Interestingly, this inadequacy is illustrated by an analogous method to the one that Loar used to try to show that that-clauses cannot adequately capture the psychological

content of a belief or other thought, i.e., by using the Fregean/Kripkean idea of a subject having distinct conceptions of the same object or state of affairs. The upshot of this is that Edwards' E-component is added to the specification of the individuation conditions of demonstrative thoughts in the following manner:

[(DEM), (E), (...is F), [t]]

In this specification, "DEM" "...stands for the psychological state undergone in thinking demonstratively of an object"; "...is F" stands for [the subject's] conception of the property whose instances are described by "...is F" (Edwards, 1994, p.161); "t" is the time of the demonstrative thought; and "E" is the E-component. The E-component is meant to contain the perceptual experience of the subject, while he or she is having the demonstrative thought, as it is this experience which may well contribute towards the individuation conditions of that particular thought. This does sound very much like what we are after in relation to individuating the belief of a subject, which is then going to be realized in a possible world. Perhaps we could use a variation of the E-component, with a conceptual role factor added to the perceptual factor; it could be called the CE-component.<sup>9</sup> The specification of the realization conditions of a belief could then be something like the following:

S believes that P iff [S believes (N)P, (CE), and REAL((N)P, that P), [t]]<sup>10</sup>

In the above specification "(N)P" would stand for the narrow belief that P, i.e., P without the anchored wide content; "(CE)" would be the CE-component, which would ensure that the narrow belief that P would also include any conceptual and



experiential elements that S was aware of in relation to this belief, and that might well help to individuate it from the subject's viewpoint; "REAL" are the realization conditions of the narrow belief (N)P in a possible world, which would result in the narrow belief being anchored to the wide content P in that world; finally, "t" is the particular time when this belief occurs. The inclusion of the CE-component will hopefully constrain the realization conditions of the narrow belief (N)P, by making sure that the original conceptual and experiential elements from S's subjective perspective also play a role in individuating the realized narrow belief (N)P, i.e., by ensuring that the wide content P that it gets anchored to when realized, will have the same subjective conceptual and perceptual properties from S's perspective that the wide content P had in S's original context.<sup>11</sup> If the above specification is deemed plausible, then it can be seen that a variation of Edwards' E-component could be a useful mechanism for constraining the possible worlds in which a belief can be realized.

To conclude this section on Stalnaker's objections to realization conditions as a candidate for narrow content, it will be useful to examine what Stalnaker sees as the "source" of the above objections, namely that the:

...narrow content is derivative from (actual and possible) wide content. The explanation of realization conditions takes for granted that somehow, the external and internal facts about a mental event or state that we have identified determine a content (in the ordinary sense) for that state. Then from the actual wide content, together with facts about what the wide content would be under various counterfactual conditions, we extract the narrow content (Stalnaker, 1990, p.200).

So, is Stalnaker correct when he says that narrow content is derivative from wide content, in relation to the notion of realization conditions? Firstly, it is fair to say that

the notion of narrow content we have been examining is derivative at least in the sense that it is an abstraction from natural language, and from the practice of explanation of human action.<sup>12</sup> However, it can be seen that Stalnaker is talking of “counterfactual conditions”, when it was noted previously that there is the possibility that Loar’s realization conditions of a belief could operate not in counterfactual conditions, but in epistemic conditions (similar to Chalmers’ notion of an epistemic intension of a thought), where the realized belief-content would be determined by, and evaluated in, a world, as if that world was the actual world. So, how does this distinction affect the question of whether narrow content is derivative on wide content? Let us go back to Bert, who we shall say has the belief that he has arthritis in his ankles. The situation is that we are not sure whether Bert has his belief on Earth (where arthritis is a disease of the joints only), or on Twin Earth (where arthritis is a disease of the joints and bones in general). Despite us not knowing this, Bert himself has the belief that he has a rheumatoid ailment in his ankles that is called “arthritis” in the social environment that surrounds him.

It seems reasonable to say that this belief of Bert’s has probably been caused by the external world, in particular, by the social and linguistic aspect of the external world. Indeed, we could say that the *intension* of Bert’s belief has been caused by interactions with the external world. By “intension” here, I mean it is a “function” from possible worlds to extensions or truth-values, depending on whether it is an intension of a concept or a thought (Chalmers, 2002, p.3). I suppose that it is roughly similar to the notion of the “intension” which is considered to determine an extension, which Putnam (1975) considered was a false assumption, certainly with regards to thoughts about natural kinds, such as water. Of course, in admitting all this, it does

not mean that I am committed to the view that the external world *constitutes* the intension or content of Bert's arthritis-belief. The external world may be causally responsible for the intension of Bert's arthritis-belief, but it could be claimed that this has been achieved by affecting an internal state of Bert's, i.e., it could be claimed that the intension of Bert's arthritis-belief is mainly a matter of how Bert is in himself (Braddon-Mitchell/Jackson, 1996, p.212; Chalmers, 2002, p.15).<sup>13</sup> If it is allowed that the intension of Bert's arthritis-belief is not constituted by his social environment, then it could plausibly be claimed that this belief-intension has a narrow content.

If we now go back to Bert and his arthritis-belief; we still don't know whether Bert is on the Earth or Twin Earth, so we don't know which social environment was causally responsible for Bert's belief. However, whichever environment it is, the intension of Bert's belief will be narrow (i.e., it could apply to either social environment), and might be something like "I have the rheumatoid ailment in my ankles called "arthritis" by the social community which I am surrounded by". Therefore, if we now consider the realization conditions of Bert's belief, which also means accepting that his belief is not a misconception, we can imagine Bert's belief on the Earth or Twin Earth being true, that is, in either world, it will be true that Bert has arthritis in his ankles. If the Earth is considered as Bert's actual world, then his arthritis-belief will be true because at that world the intension of his belief is effectively determining the extension or content of that thought. The intension of his "arthritis" concept will determine that the extension of that concept is what "arthritis" means on the Earth, i.e., a rheumatoid ailment that affects the joints only. So Bert's use of the term "arthritis" will be semantically correct in relation to the Earth, which means that the intension of Bert's arthritis-belief as a whole, will be true. If Twin Earth is

considered as Bert's actual world, then it will be the same story as it was on the Earth, the only difference being that the extension that Bert's arthritis concept determines in this world will be "a rheumatoid ailment of the joints and bones in general". I'm not sure how convincing all of this is, but if accepted, it might show that the narrow content defined by realization conditions is not necessarily derivative on wide content.

Indeed, it appears that Bert's narrow-belief is actually determining what the wide content of the belief will be in either world.<sup>14</sup> As previously stated, the reason for this seems to be that Stalnaker considers counterfactual worlds when evaluating the content of a thought. He is quite right when he says that realization conditions involve assuming that the external and internal facts about a mental state are taken to determine a particular content, but Stalnaker then claims that the actual wide content for the thought, along with various counterfactual wide contents, plays a part in "extracting" the narrow content, which I don't think is the case. With Loar's realization conditions, we don't know which world is going to be actual until it is selected, and where we then evaluate the thought-content relative to that world. I can understand why Stalnaker makes his case as he does, as it does seem that if you look at the actual and various counterfactual wide contents for an arthritis-belief, then some "narrow content" that is in common between all of them will probably emerge. However, as I have hopefully shown, it is actually the opposite which is the case with realization conditions, as the narrow content or intension of Bert's arthritis-belief can be formed in a suitably internalistic manner, which does not depend on the wide content of the belief.

Therefore, this chapter has been concerned with explicating and assessing Loar's notion of context-indeterminate realization conditions, which he hopes will provide the narrow psychological content of beliefs with intentional properties. The criticisms of Stalnaker were also examined and hopefully deflected.

### 3.3 PART 1 CONCLUSION

This part of the thesis has been concerned with a conceptual role construal of narrow content; and has focussed mainly on Loar's theory of psychological and social content. A conceptual role theory of thought-individuation usually forms the narrow component in so-called two-factor theories of content, with the wide component of thought-individuation usually being a referential truth conditional theory of content. With Loar, the psychological content of a belief is the narrow component and the social content of a belief is the wide component. I thought that Loar argued persuasively that just because two de dicto belief ascriptions have the same obliquely occurring general terms in them, it is not necessarily the case that the two beliefs also have the same psychological content. For Loar, the psychological content of a belief consists in the way that the belief is individuated by its conceptual relations to the subject's sensory inputs, other beliefs/attitudes and behavioural output. When applied to Burge's "arthritis" thought experiment, it was again persuasively argued that while the social content of actual-Bert and the doctor's *arthritis* concepts are the same, the psychological content of those *arthritis* concepts are different. The other point is that whilst the social content of actual-Bert's and counterfactual-Bert's *arthritis* concepts

seems to be different, it seems intuitive to consider the psychological content of their *arthritis* concepts as the same.

The main problem for Loar's theory, and indeed, for the conceptual role individuation of beliefs generally, is that it is difficult to actually specify what the content of the particular narrow belief is. With the wide content of a belief it is fairly straightforward to specify this, as it usually turns out to be the conventional meaning for the term that is contained in the belief. However, with conceptual role individuation, the content of a belief is determined by its inferential relations with other beliefs and the resultant behaviour of the subject, which makes it difficult to give a determinate specification of the belief's content.

This was a tricky situation, and I'm not sure I ever got to the bottom of it, but I did try out some strategies to find an answer to this problem. Firstly, I pondered the use of subjective and objective perspectives on ascribing beliefs. When utilising the subjective viewpoint of the subject, all this entails is using the disquotational principle of belief ascription, i.e., if the subject sincerely assents to a proposition  $p$  that is put to him/her, then one would say that the subject believes that  $p$ . The objective viewpoint on ascribing beliefs entails that the conventional concept or meaning of the term involved with the belief would play a role in individuating that belief.

Consideration was also given as to whether de re or de dicto belief ascriptions can be used, perhaps in conjunction with the subjective and objective perspectives, in specifying the narrow psychological content of a subject's belief. This seems a non-starter with de re beliefs, whether one uses the subjective or objective perspective

when ascribing the belief, as with de re beliefs it is the specific referent of the belief in a particular context which plays the main role in individuating that belief. The upshot of this is that with a belief that contained a term like “arthritis”, in the actual situation this would mean that “arthritis” has the content “inflammation of the joints only”, whilst in the counterfactual situation, the Twin-English term “arthritis” would have the content “inflammation of the joints and bones”. This means that the term “arthritis” cannot be used in a narrow belief ascription that is shared by actual-Bert and counterfactual-Bert.

A more promising approach for specifying the narrow psychological content shared between actual-Bert and counterfactual-Bert is to provide a de dicto ascription of their beliefs. In contrast to a de re reading, a de dicto reading of a belief is meant to provide and characterise the content of that belief for the subject, to capture how the subject conceives of things with the belief. However, even on a de dicto reading, it is problematic to specify a shared narrow psychological content between two beliefs. The reason for this is that the normal understanding of de dicto belief ascriptions seems to entail that general terms contained in those beliefs have the conventional meanings or concepts attached to them. The upshot of this is that in the actual situation the term “arthritis” has the concept of being an inflammation of the joints only, whilst in the counterfactual situation the Twin term “arthritis” has the concept of being an inflammation of the joints and bones. Therefore, as with the de re reading, it seems that even on a de dicto reading, we cannot specify what the shared psychological content is between the Berts’ arthritis beliefs.

It was then considered whether perhaps the idea of subjective and objective viewpoints could be utilised with de dicto belief ascriptions. We can say that with an objective de dicto belief ascription, the general term involved in the belief has the conventional concept attached to it. Then we can say that with the subjective de dicto belief ascription, the general term involved in the belief has the subject's idiosyncratic concept attached to it. However, to specify what the idiosyncratic concept attached to, say, actual-Bert's "arthritis" term is, seemed to be arbitrary, as actual-Bert might well attach any concept to his "arthritis" term. Nevertheless, it seems plausible to me that the narrow *arthritis* concept that is shared by actual-Bert and counterfactual-Bert could be a more general, descriptive, concept, something like "the rheumatoid disease in my social environment".

Therefore, the crucial issue is how Loar's conception of psychological content fares in satisfying the three conditions of adequacy for narrow content. It seems to satisfy condition (1), as the psychological content of a belief is understood to be an internal, "in the head", property of the subject. However, in its original form, it does not seem to satisfy conditions (2) and (3). That is, the psychological content of a belief cannot be given a specification using a that-clause, which in turn means that a psychological generalisation which subsumes the behaviour of the duplicates cannot be given. My idea of a subjective de dicto belief ascription was an attempt to remedy this, but I still have my doubts about whether Loar's psychological content, and the conceptual role individuation of beliefs in general, can provide an adequate narrow content.



## Notes to Chapter 3

<sup>1</sup> It is perhaps worth noting here that a conceptual role specification of a belief could still involve “internal” truth conditions, i.e., “the state caused by state x and which leads to state y...etc.).

<sup>2</sup> Moreover, the person’s conception of how things are could well inform their consequent actions, which could give further support to the idea that their narrow belief does have content.

<sup>3</sup> Perhaps this situation is analogous to Nagel’s (1974) view of subjective and objective perspectives on the phenomenon of consciousness. Loar (1988b) makes a similar point on p.135.

<sup>4</sup> Something similar to Loar’s notion of realization conditions is also used by Jackson/Pettit (1993) to give truth-evaluability to what they call narrow “predictive content”, which is understood to be “...sets of possibilities associated with a person’s beliefs and desires” and that “...content in terms of sets of possibilities is automatically truth-evaluable. It comes out true precisely if the content set of possibilities contains the way things actually are” (1993, p.269).

<sup>5</sup> It is worth noting that the views of Chalmers (2002, 2003a) on narrow content also have some similarities with Loar’s realization conditions and Stalnaker’s possible worlds portrayal of them. Essentially, Chalmers believes that the content of a thought, such as a belief etc., can be broken down into two components, a *subjunctive* content and an *epistemic* content (2002, p.2). These two contents approximately compare to the above two methods of Stalnaker that we used to determine and evaluate the arthritis beliefs of Bert on various possible worlds. The subjunctive content of a thought is equivalent to the content that a thought has when it’s actual world content is fixed, and then this fixed-content thought is evaluated at counterfactual worlds. In contrast, the epistemic content of a thought is equivalent to the content that a thought has when said content is determined at the same world that it is going to be evaluated in, i.e., this is not a counterfactual evaluation, as the content of the thought has not been fixed by the content that it has in the actual world.

It is this epistemic content of a thought that Chalmers views as being a kind of narrow content. So, for example, the subjunctive content of a water-thought on Earth would be H<sub>2</sub>O, while the subjunctive content of the same water-thought on Twin Earth would also be H<sub>2</sub>O. The epistemic content of a water-thought on Earth would be H<sub>2</sub>O, however, the epistemic content of the same water-thought on Twin Earth would be XYZ (2002, p.3). With the epistemic content of a thought, the content that it is determined to have at a particular world, is considered to be its actual content, so its epistemic content could vary several times at several different worlds. The subjunctive content of a thought is fixed once at the actual world, and stays the same at every other world. Moreover, Chalmers believes that the intension of a thought or concept can also be subdivided into an *epistemic* intension and a *subjunctive* intension, with the epistemic intension picking out the epistemic content at a world and the subjunctive intension picking out the subjunctive content at a world.

<sup>6</sup> Yet another way to put this, this time in Kaplanesque terminology, is that in regard to the meaning of the linguistic term “arthritis”, Bert will be a “consumer”, who uses that term with a “pre-packaged” meaning that has been created by the social community he is surrounded by. This way of putting things is from Segal (2000) p.62.

<sup>7</sup> It will be recalled, that Loar does in fact declare that his notion of realization conditions does include a “...combination of indexicality and substantive conceptual content” (Loar, 1988a, p.197). It was because of their different conceptual roles that Loar believed his conditions would narrowly individuate “water” and “alcohol” beliefs as distinct, in contrast to Fodor’s (1987) notion of narrow content.

<sup>8</sup> Chalmers, 2002, p.26 seems to suggest a similar idea to that of Edwards’ notion of an E-component, where he suggests that a centred world might require a built-in “marked experience” to cope with the content of perceptual demonstratives.

<sup>9</sup> Perhaps I should call it the C of E-component, as I think I’m going to need divine assistance to boost the plausibility of what follows.

<sup>10</sup> It should be noted here that as well as enlisting the help of Edwards in this specification, I have also shamelessly cribbed from Stalnaker, 1999, p.24, who himself was influenced by Field (1978).

<sup>11</sup> It is the *subjective* conceptual role properties that we want to be the same, as the *objective* conceptual properties of P in the realized context may well be different, but only different within reason, due to the subjective constraints. That is, in the actual world the objective conceptual properties of “arthritis” is that it is an inflammation of the joints only, but in another possible world the objective conceptual properties of “arthritis” could be that it was an inflammation of the joints and bones. The objective conceptual properties are different, but not radically different, such as “arthritis” being blue furry toys.

<sup>12</sup> This was pointed out to me by my supervisor.

<sup>13</sup> This is of course the starting-point for the thesis of “methodological solipsism”, which roughly says that the content of a thought is not logically dependent on what caused it, or on the external world in

general. The term “methodological solipsism” was first coined by Putnam (1975), and later discussed in some detail by Fodor (1980).

<sup>14</sup> Chalmers, 2002, p.26 makes a similar point in relation to his epistemic intension/content.

## **PART 2**

# **INDEXICAL NARROW CONTENT**

In this part of the thesis an indexical construal of narrow content will be examined, the exemplar of which is Fodor's so-called "mapping theory". In chapter 4, the mapping theory will be explicated by following Fodor's reasoning and motives as he moves towards his postulation of an indexical narrow content. In chapter 5 the mapping theory will be assessed and criticised, with a particularly searching critique coming from Block (1991), who, among other things, accuses the mapping theory of collapsing into syntax.

## CHAPTER 4

### FODOR'S "MAPPING THEORY"

Fodor's (1987) indexical conception of narrow content will now be examined. His conception is slightly confusing, so I won't explain the indexical element that it contains immediately. My strategy will be to follow the path that Fodor cuts through the philosophical undergrowth, which ultimately leads to his notion of indexical narrow content.<sup>1</sup>

#### 4.1 Methodological Individualism and Causal Powers

One of Fodor's aims is to argue for the thesis of "methodological individualism", viz., that a scientific psychology only needs to individuate mental states "...with respect to their causal powers" (Fodor, 1987, p.42). Fodor is happy to acknowledge that the commonsense way of individuating mental states or propositional attitudes is relational, but he does this on the condition that it is understood that "...no property of mental states, relational or otherwise, counts taxonomically unless it affects causal powers" (Fodor, 1987, p.42). Indeed, Fodor gives an argument that is supposed to show how even though the content of mental states may differ in their relational properties, they would still be considered as identical if one was individuating by the causal powers of those states. Fodor's strategy is of interest here, as he brings in

cross-context comparisons and counterfactuals to argue his case, which is what he also does in his later argument for narrow content.

To examine Fodor's "cross-context test" for the identity of causal powers, imagine Putnam's classic Twin Earth scenario; while I am on Earth, my doppelganger, Twin-Me, is on Twin Earth. Being my doppelganger means that Twin-Me is a "molecule for molecule" duplicate of myself. As is usual, Twin Earth is physically, phenomenally, and linguistically identical to Earth, except for one thing; where water on Earth is composed of H<sub>2</sub>O molecules, "water" on Twin Earth is composed of XYZ molecules. Having set up this scenario, now imagine that I think and utter the phrase "Bring water" and that Twin-Me also thinks and utters the phrase "Bring water". The Putnamian analysis of this situation will be that even though me and Twin-Me are molecular duplicates, that we have identical brain states, etc., the content of our respective thoughts and utterances will be different. And this difference depends on a relational property, that is, my thoughts/utterances will concern H<sub>2</sub>O, while Twin-Me's thoughts/utterances will concern XYZ. So, when I say "Bring water" I will be brought H<sub>2</sub>O, while Twin-Me's utterance of "Bring water" will result in him being brought XYZ (Fodor, 1987, p.35). However, Fodor views this difference as:

...irrelevant to the question about identity of causal powers, *because these utterances (/thoughts) are being imagined to occur in different contexts* (mine occur in a context in which the local potable is H<sub>2</sub>O, his occur in a context in which the local potable is XYZ). What is relevant to the question of identity of causal powers is the following pair of counterfactuals: (a) If his utterance (/thought) had occurred in my context, it *would have had* the effects that my utterance (/thought) did have; and (b) if my utterance (/thought) had occurred in his context, it *would have had* the effects that his utterance (/thought) did have (Fodor, 1987, p.35).

Fodor goes on to say that for the causal powers of these thoughts/utterances to be the same, both of the above counterfactuals have to be true. And, initially, it does seem plausible to say that both counterfactuals would indeed be true. If Twin-Me had been in my context, then his thought/utterance “Bring water” would have had the same effect as my thought/utterance of “Bring water” in that context, i.e., we would have both been brought H<sub>2</sub>O. Alternatively, if I had been in Twin-Me’s context, my thought/utterance “Bring water” would have had the same effect as his thought/utterance “Bring water” in that context, i.e., we would have both been brought XYZ. Fodor takes this to show that despite relational differences, a scientific psychology which individuated mental states by causal powers, would view the respective thought-contents of me and Twin-me as identical. This also ties in with another of Fodor’s aims, that of preserving the supervenience of mental states on neurophysiology. If mental states are individuated with regard to their causal powers, this would lend itself to the view that causally efficacious mental states need to be locally supervenient on brain states. A corollary of this for Fodor is that identical brain states would have identical causal powers.

#### 4.1.1 *Possible Problems for Fodor’s Cross-Context Test*

However, are things as straightforward here as Fodor claims they are? For example, an externalist could say that Fodor is begging the question here, in that the causal powers of me and Twin-me do differ because “Bring water” thought and uttered on Earth results in H<sub>2</sub>O being brought, whilst “Bring water” thought and uttered on Twin Earth results in XYZ being brought.<sup>2</sup> That is, in the situation of the standard Twin

Earth thought experiment, where I am on Earth and Twin-me is on Twin Earth, it is still plausible to claim that the causal powers of our respective water thoughts are different. Moreover, this difference in our causal powers would depend on differences in our relational properties, e.g., differences in our respective physical environments.

Indeed, there is the possibility that the above externalist position could be taken further, to show that the causal powers of me and my twin's "Bring water" are not the same even in the same context. If I am on Twin Earth and utter "Bring water", my request is not satisfied, as Twin-water will be brought to me, and vice versa for Twin-me on Earth.<sup>3</sup> A similar point is made by Edwards (1994) who utilises an idea of Burge (1989), where the latter imagines that there is a "device" that is capable of determining peoples' causal histories (Edwards, 1994, p.34). Edwards then considers the use of such a device with the above cross-context situation of Fodor's. It seems plausible that if, for example, I am on Twin Earth and think/utter "Bring water", the device would know from my causal history that what I am asking for is water (H<sub>2</sub>O) and not Twin-water (XYZ), and so water would be brought to me instead of Twin-water, and vice versa for Twin-me on Earth (Edwards, 1994, p.34). Edwards claims that this "...indicates that even if The Context Claim (cross-context test) is accepted the circumstances of a test in which the causal powers of the Twins differ can be described" (Edwards, 1994, p.34).

There are several threads involved with the above externalist counter-arguments, such as Kripke's and Putnam's view of natural kind terms and the causal-historical theory of reference. When I utter "Bring water" on Twin Earth, because it is H<sub>2</sub>O that I have

been referring to with my “water” terms on Earth, then that is what I am asking for on Twin Earth, not Twin-water which is composed of XYZ, or so the story goes.

However, it still seems to me that there is one reading of Fodor’s “cross-context” claims that does seem plausibly correct, viz., if me and my twin utter “Bring water” on Twin Earth, then our utterances/thoughts do have the same effects in that we are both brought XYZ, and the same, *mutatis mutandis*, for my twin and I on Earth. That is, in the same contexts, me and my twin’s thoughts/utterances do seem to have the same effects, and so, also seem to have the same causal powers. The plausibility of this stems from the fact that, at least in the standard Twin Earth thought experiments, “water” on Twin Earth is XYZ and “water” on Earth is H<sub>2</sub>O. Even though I might know that on Twin Earth when I think/utter “Bring water” what I intend to be brought is H<sub>2</sub>O, this doesn’t alter the fact that the causal effect of my utterance will be the same as Twin-me’s, i.e., XYZ will be brought to both of us.

Therefore, it seems to me that with regards to the individuation of psychological states, two states having the same consequences in the same contexts, is a result which plausibly points to them having identical causal powers, and so being type-identical as far as a scientific psychology is concerned.<sup>4</sup> Having said this, I am not so sure about Fodor’s more general claim that science taxonomises according to causal powers.

There may be some supportive examples in the physical sciences. For example, from chemistry, if substance A and substance B have identical causal effects from identical chemical processes in the same contexts, then it might be plausible to taxonomise those substances as the same, due to their identical causal powers. However it seems less plausible to claim that this is the case with the physical sciences as a whole.



Nevertheless, just as it could be claimed by the externalist that the cross-context test for identity of causal powers of mental states is question-begging, Fodor's claim about the local supervenience of mental states on neurophysiology could also be viewed as question-begging. If the supervening properties of a person's mental states are considered to be internal or "in-the-body" ones only, then it might be plausible to claim that they would also only be locally supervenient on neurophysiology. However, if the supervening properties are considered to include ones that are relational or beyond the body, then Fodor's claim of local supervenience of mental states on neurophysiology may falter. That is, the supervenience base of a person's mental states would include the person's environment as well as their neurophysiology.<sup>5</sup>

#### 4.2 Does Identity of Thought-Content Determine Identity of Extension?

However, even if we allow Fodor's cross-context claim about the identity of causal powers to go through, it seems that he is still concerned with the effect that the relational differences seem to have on thought-contents, which Putnam's Twin Earth scenario seems to highlight. In particular, it seems that:

*...the identity of the contents of mental states does not ensure the identity of their extensions: my thoughts and my Twin's...differ in their truth conditions, so its an accident if they happen to have the same truth values (Fodor, 1987, p.45).*

So, Fodor's problem is that even if one agrees that respective thought-contents are identical, as individuated according to the identity of their causal powers, the thoughts

could still have different truth-conditions, as what they are referring to, their extensions, are not the same. Indeed, this point was seen by Putnam (1975) as one of the major lessons to come out of his Twin Earth example. It had been an influential view in the philosophy of mind that an internal thought-content (or intension) determined the extension of that thought.<sup>6</sup> So, if two subjects had identical thought-contents, then it was taken that the extensions of their thought-contents were also identical, that is, they were both referring to the same object or state of affairs. This also meant that the truth conditions and subsequent truth values of the subjects' contents were also the same. For example, if two subjects had the same thought in their heads, such as, "Water is wet", the extensions of these respective thoughts would be the same, they would both refer to the natural kind "water" (which is composed of H<sub>2</sub>O) in the subjects' environments. Moreover, the truth conditions of the respective thoughts would also be the same, both thoughts would be either true or false together, depending on whether *water was actually wet in that world*. This comparison of a thought-content and the relationship it bears to the world, in order to produce a truth value for the thought, is called "semantic evaluation" by Fodor (Fodor, 1987, pp.10-11).

However, what Putnam's Twin Earth seems to show is that two neurophysiologically identical subjects could appear to have identical thought-contents of something in the world, but that the extensions of those thoughts are different, and so the truth conditions are also different. What this was supposed to show is that an internal thought-content did not entirely determine its extension, i.e., that the identity conditions of the thought necessarily make reference to features outside the head.

As Fodor concedes, this result was problematic for philosophical psychology, as the attribution and individuation of thought-contents is usually done using the 2-way relationship between content identity and extensional identity. Basically, it was thought that a content determined an extension, in that there would be no change of extension without a change of content. Conversely, the extension of a thought constrained what the content of that thought could be, and if there was a change in extensional identity of a thought, then there was a corresponding change in the content identity of that thought (Fodor, 1987, p.46). However, Putnam seemed to have called this methodological presupposition into question; as Fodor puts it “The Twin Earth Problem *is a problem, because it breaks the connection between extensional identity and content identity*” (Fodor, 1987, p.47). Therefore, it seems that Fodor has two problems. (1) Even though it might be said that two thoughts are content-identical, as judged by their causal powers, Fodor would also like the truth-conditions of those thoughts to be the same as well, i.e., being either true or false together. (2) He would like to repair the connection between extensional identity and content identity.

#### 4.2.1 *How Fodor Repairs the Connection Between the Content Identity and Extensional Identity of Thoughts*

Fodor’s solution to these problems, which also includes his conception of narrow content, is where the indexical element enters the story. The solution is not original (as Fodor himself acknowledges), it is based on the ideas of White (1982), which in turn were based on Kaplan’s notion of the character of an indexical term, where the

character acted like a function from contexts of utterance to contents.<sup>7</sup> Fodor's

solution to the above two problems is the following:

*The Twin-Earth examples don't break the connection between content and extension: they just relativize it to context...extensional identity still constrains intentional identity because contents still determine extensions relative to a context. If you like, contents are functions from contexts and thoughts onto truth conditions (Fodor, 1987, p.47).*

To illustrate this point, imagine the Twin Earth scenario again, I am on the Earth and my doppelganger, Twin-Me, is on Twin Earth. It seems plausible to say that in our respective contexts, the contents of our respective thoughts still determine extensions. In my "Earth-context", say, the content my "water"-thought still correctly determines its extension, which is H<sub>2</sub>O on Earth. Meanwhile, in Twin-Me's "Twin Earth-context", the content of his "water"-thought also correctly determines its extension, which is XYZ on Twin Earth. Therefore, my "water"-thoughts on Earth will always be about H<sub>2</sub>O, while Twin-Me's "water"-thoughts on Twin Earth will always be about XYZ.

This also means that in the context of Earth, the content of my "water"-thoughts, i.e., H<sub>2</sub>O, acts as a function onto truth conditions regarding my "water"- thoughts, i.e., whether the thoughts are true is dependent on my being in the context where water is H<sub>2</sub>O. With regards to the context of Twin Earth, the content of Twin-Me's "water"- thoughts, i.e., XYZ, acts as a function onto truth conditions regarding his "water"- thoughts, i.e., whether the thoughts are true is dependent on his being in the context where "water" is XYZ. So, even though the extensions of me and Twin-me's "water" term are context-sensitive, in that it picks out H<sub>2</sub>O in the Earth-context and XYZ in the Twin Earth-context, it can still be claimed that there is a narrow "character"

shared between our water thoughts which enables both our “water” terms to have the same extensions in the same contexts.

Furthermore (following Fodor), we can call the relationship between Twin-Me and Twin Earth which results in his “water”-thoughts being about XYZ, condition C; while my relationship with Earth which results in my “water”-thoughts being about H<sub>2</sub>O can be called condition C’ (Fodor, 1987, p.48). Fodor then makes the following claim:

Short of a miracle, it must be true that if an organism shares the neurophysical constitution of my Twin *and satisfies C*, it follows that its thoughts and my Twin’s thoughts share their truth conditions (Fodor, 1987, p.48).

And of course, by hypothesis, the organism which shares the neurophysical identity with my Twin, is myself. This then allows Fodor to propose the following counterfactual: If I am in my Twin’s context, then given the neurophysical identity between us, my “water”-thoughts will be about XYZ iff his are (Fodor, 1987, p.48). Fodor claims that this counterfactual is necessarily true, and it does intuitively seem that this is the case. Just by being in my Twin’s context, I seem to satisfy C, as it is because of this particular condition that my Twin’s “water”-thoughts always denote XYZ; that is, due to the fact that the watery stuff in my Twin’s environment is actually composed of XYZ, this is the substance that he has been causally interacting with and which is the extension of his “water”-thoughts. Therefore, according to Fodor, once I enter his context, the extension of my “water” thoughts will also be XYZ, as this is the substance that I will then be causally interacting with. This also means that my “water”-thought and my Twin’s “water”-thought will be sharing their truth conditions, as they will both be concerning facts about XYZ. And of course, it

will be the same story if my Twin comes into my context, the only difference being that the substance our respective “water”-thoughts will both be denoting is H<sub>2</sub>O.

It can be seen that the above idea is similar to Fodor’s cross-context test for the identity of causal powers which was examined earlier. Moreover, as there, the problem raised by the causal-historical theory of reference could once again be raised here. In Putnam’s original Twin Earth thought experiment, when Earth’s Oscar travels to Twin Earth, sees some XYZ in a glass, and thinks “water is wet here”, the usual analysis is that the liquid Oscar is referring to is not “water” as Oscar understands it. Because on Earth Oscar has been in a causal-historical relationship with H<sub>2</sub>O, which is what water is composed of on Earth, the expression “water” which he uses is (to use a term of Fodor’s) “anchored” to H<sub>2</sub>O. Oscar’s expression “water” has its reference fixed to H<sub>2</sub>O, because this is what it has always referred to in Oscar’s Earth-context (the term “water” is a “rigid designator”, as Kripke would say).

However, does this mean that if Oscar has a “water”-thought, the *content* of that thought is also similarly fixed or anchored to H<sub>2</sub>O? Fodor seems to be arguing that it does not. When Oscar goes to Twin Earth, even though his expression “water” cannot be correctly used to refer to XYZ, the content of his “water”-thought will now be individuated by the substance he is causally interacting with, which will be XYZ on Twin Earth. Therefore, once we have a situation where the substance being referred to by the Twins is identical, i.e., the extensions of their thought-contents are identical, then it will follow that the truth conditions of their respective thought-contents will also be the same. So, Fodor at least is satisfied that he has achieved his aims of

restoring the link between content identity and extensional identity and of having the Twins' thoughts sharing their truth conditions.

#### 4.3 The Extensional Identity Criterion for Thought-Contents

However, Fodor is also convinced that he has something else as well, he claims:

But now we have an extensional identity criterion for mental contents: Two thought contents are identical only if they effect the same mapping of thoughts and contexts onto truth conditions (Fodor, 1987, p.48).

By this, Fodor means that my thought-content is identical to Twin-Me's thought-content, only if in every context where my thought has a certain truth condition, Twin-Me's thought has the same truth condition, and vice versa (Fodor, 1987, p.48).

By this criterion, Fodor is able to claim that the content of my "water"-thought is identical to the content of Twin-Me's "water"-thought, as when we are both in either of each other's contexts, it means that our respective thoughts will have the same truth conditions, as both thoughts will be referring together to either XYZ or H<sub>2</sub>O. This is despite the fact that when we are each in our own contexts, that is, Earth and Twin Earth, our respective thoughts will be referring to different substances and so have different truth conditions (Fodor, 1987, p.48).

So, as mentioned above, Fodor's contextual argument for content identity is similar to the contextual argument he earlier used to show how the causal powers of the Twins' respective thoughts/utterances were also identical. There it was claimed by Fodor that my thought/utterance "Bring water" would have the same effect as Twin-Me's

thought/utterance “Bring water”, when we were both in the same context; i.e., if we were both on Earth it would result in H<sub>2</sub>O being brought to us, while if we were both on Twin Earth it would result in XYZ being brought to us. This was despite the fact that when we were each in our own contexts, the causal powers of our respective thoughts/utterances might well have been judged to have different effects, i.e., H<sub>2</sub>O being brought to me and XYZ being brought to Twin-Me.

So, in the one case, it is the identical effects which are used to infer the identical causal powers of thoughts, and in the other case, it is the identical truth conditions which are used to infer the identical thought-contents. Fodor’s strategy, then, appears to be that when judging the identity or otherwise of various properties of thoughts, this should be done with the thoughts being in the same context, rather than each being in a different context.

#### 4.3.1 *The Extensional Identity Criterion for Narrow Content*

Indeed, Fodor claims that his extensional identity criterion for mental contents is also an extensional criterion for narrow content itself (Fodor, 1987, p.48). So, what exactly is Fodor’s conception of narrow content? Well, it turns out to be something which has already been mentioned, i.e., it is the thought-content that acts like a function from thoughts and contexts onto truth conditions. As Fodor puts it:

...a narrow content is *essentially* a function from contexts onto truth conditions; different functions from contexts onto truth conditions are ipso facto different narrow contents (Fodor, 1987, p.53).



And, from what has gone before, we can infer that the narrow content that two thoughts share is that function which results in the same mapping of thoughts and contexts onto truth conditions. So, when Twin-Me counterfactually enters my Earth-context, and we then both think that water is wet, according to Fodor, our thoughts will be content-identical, because given that we are both in the same context, our thoughts will have the same truth conditions. As explained above, the reason for the same truth conditions is that both Twin-Me and I will be referring to the same substance, i.e., on the Earth it will be H<sub>2</sub>O; so the content of our respective “water”-thoughts, i.e., *that water is wet*, will also be the same, so leading to the same truth conditions for both thoughts. It would be the same story if I counterfactually entered Twin-Me’s Twin Earth context, the only difference being of course that we would both be referring to XYZ in that context.<sup>8</sup>

From both of these counterfactuals, it can be seen that the function or content that me and my Twin’s thoughts both share, which effects the same mapping of thoughts and contexts onto truth conditions, is that *water is wet*. Therefore, I think that Fodor would say that the narrow content between myself and my Twin would be something like “the thought that water is wet” (Fodor, 1987, p.51). Why Fodor claims that the above content-identity criterion is also an “extensional criterion for narrow content” is presumably connected with him also claiming above that a different function from contexts onto truth conditions would mean a different narrow content. By this I think he means that the extension of the two thoughts being counterfactually compared would constrain the function of the thoughts in question. For example, if me and Twin-Me were both thinking that gold is yellow in the same context, the function that our thoughts would share, which would map our thoughts and contexts onto truth

conditions, would be something like “the thought that gold is yellow”. This would be a different narrow content from “the thought that water is wet” due to the extension of the thoughts being different.<sup>9</sup>

#### 4.4 How Does Fodor Specify his Narrow Content?

There are a number of issues which should be mentioned in connection with Fodor’s conception of narrow content presented here. For example, Fodor has the problem of not being able to express what the narrow content is that my Twin and I share. It might be thought that this is a strange thing to say, as we seem to have just expressed quite easily, that the narrow content that my Twin and I share is “the thought that water is wet”. It is here that Fodor draws a distinction between “expressing” the narrow content and “abstracting a specification” of that content; claiming that the former is not possible while the latter is. The reason that one cannot “express” the narrow content “that water is wet”, is because “water” is a term that is “anchored” to the context of its use. By saying that a term is anchored, Fodor means that “it has a determinate semantic value” (Fodor, 1987, p.49).

So, in the Earth-context, the English term “water” is anchored to H<sub>2</sub>O, i.e., it always refers to H<sub>2</sub>O; while in the Twin Earth-context, the Twin English term “water” is anchored to XYZ, i.e., it always refers to XYZ. It is because of this anchoring of the term “water”, that “that water is wet” cannot be used to express what content my Twin and I have in common, because for me it will mean that H<sub>2</sub>O is wet, while for Twin-Me it will mean that XYZ is wet. Indeed, it is because of this that Fodor claims

that his narrow content is “only content *potentially*” (Fodor, 1987, p.50), until it becomes anchored to a context, and is then capable of semantic evaluation relative to that context. By “semantic evaluation” Fodor means that a thought is judged to be true or false by the relation it has with the world, and in particular, its relation to the fact or state of affairs that the thought itself is about.

So, according to Fodor, the narrow content that me and my Twin share, relating to the wetness of water, has no “real” or “actual” content, until me and my Twin are both located in either one’s context, where our thoughts are semantically evaluable, and the narrow content can be picked out via the same truth conditions both our thoughts have. However, even here, this picking out of narrow content cannot involve *expressing* the narrow content that my Twin and I share. Even if, for example, my Twin was in my context, and we agreed that my twin’s term “water” could now get anchored to H<sub>2</sub>O, which is what my term “water” would be anchored to, the narrow content expressed here by saying “that water is wet” would be equivalent to saying that the narrow content between my twin and I is “that H<sub>2</sub>O is wet”. And of course, this would not be true for my Twin, as he is only counterfactually in my context, he is really back on Twin Earth, where his thinking that water is wet would certainly not mean that he thought H<sub>2</sub>O was wet.

(It is plausible that Fodor would only have this difficulty of expressibility of narrow content with terms which had anchored content or were semantically determinate, e.g., natural kind terms such as “water”. If a term were deemed not to be anchored or to be “semantically neutral”, it could be easier to actually express the narrow content that two thoughts share. Indeed, if Twin Earth was physically indiscernible to the

Earth (with the exception of the “water”), the narrow contents of many thoughts could easily be read off without the aid of Fodor’s contextual arguments. For example, if me on Earth and my Twin on Twin Earth both think that those men are bald, then it would seem that the narrow content between our respective thoughts could be expressed as “those men are bald”. In fact, it seems that the wide content of our thoughts would also be the same, as they would have the same truth conditions, i.e., that the men were bald. At least I think this is correct, unless some sort of Burge-type conceptual difference could be made between my term “bald” and my Twin’s term “bald”?)

It is because of considerations like these, that Fodor takes the view that the narrow content which is shared by my Twin and I, can only be a content that is not anchored and not capable of semantic evaluation. Once a narrow content does become anchored in a context, and so is capable of semantic evaluation, then it becomes wide content (Fodor, 1987, p.48). It seems that the semantic evaluation of a thought, i.e., whether it is true or false, is an “externalising” factor in relation to mental content, as aspects of the environment are brought in to judge whether a thought-content accurately represents a situation or state of affairs in the world. But what is the relationship between a thought-content and the truth condition it is said to have? If a thought-content is judged to be false, does that mean it is false that the thought ever had that content, or rather, that the thought had that content, but that it is false in relation to the world?<sup>10</sup> Indeed, if it is the former, then friends of narrow content would do well to avoid the whole issue of truth conditions for thoughts, as they do seem to “externalise” the contents that they have. At least with the latter interpretation of truth conditions, there appears to be a space for some sort of narrow

content, something along the lines of how the subject perceives and conceives the world, which might be falsified by the actual situation in the world, but which would still count as the content that their thought, attitude, intentional state, etc., did have. This idea seems to bring in the distinction between appearance and reality in relation to the external world. On our respective worlds, my Twin and I might have identical “narrow”, subjective, perceptions and conceptions of the world we are living in, but the “wide” actuality of what we are perceiving and conceiving may differ (“water” may be H<sub>2</sub>O or XYZ depending on which context we are in).

However, to recap; Fodor does not think it is possible to *express* the narrow content that the water-thoughts of my Twin and I would counterfactually share with each other, i.e., “that water is wet”. It is now that Fodor suggests a way of “sneaking up” on the narrow content, by “mentioning” the sentence that contains the words, as an *abstraction* from “the consequences of its being anchored” (Fodor, 1987, pp.50-51). He does this, as mentioned above, by putting the narrow content that my Twin and I share in the form of “the thought that water is wet”. Presumably, because Fodor is going away from anchored content and so semantically evaluable content, “water” in this abstraction is taken to mean something like the “transparent, colourless, odourless, liquid that I perceive and conceive in my environment”. A definite description of “water” such as this would not be seen as fixed or rigid, but one that could “float” above the anchoring of terms that goes on in contexts.<sup>11</sup> Of course, this might be taken to mean that the contents of my Twin and I’s “water”-thoughts did not determine the extensions of those thoughts. But I think this would only be the case with Fodor’s *abstraction* of the narrow content that my Twin and I share. In the counterfactual situation of my Twin and I being in the same context together, then the

contents of our water-thoughts would determine their extensions and the thoughts would have the same truth conditions, i.e., relating to facts about either H<sub>2</sub>O or XYZ.

Indeed, it might be said that the main thrust of Fodor's enterprise here has been to ensure that thought contents (or intensions) do determine extensions, at least relative to a context, and that the truth conditions for both my and my Twin's thoughts are the same. Perhaps the idea of an extensional criterion for narrow content occurred to Fodor as he was working on these other aims, or perhaps not.<sup>12</sup> After all, we could have got an abstraction of narrow content from anchored content, such as "the thought that water is wet", without Fodor's ideas about contexts. One could just imagine me on Earth and Twin-me on Twin Earth both thinking that water is wet, and then abstract away a form of words that were not anchored or semantically evaluable. It seems intuitively plausible that an abstraction of narrow content such as "the thought that water is wet" could be got from an example such as this.

But perhaps I am being unfair to Fodor, as he could claim that his extensional criterion which determinately fixes the narrow content that my Twin and I share, as "the thought that water is wet" is the "unique" thought "that yields the truth condition *H<sub>2</sub>O is wet* when anchored to my context and the truth condition *XYZ is wet* when anchored to his" (Fodor, 1987, p.51). I suppose that the plausibility and coherence of Fodor's extensional criterion for mental content and narrow content will depend to a large extent on how one feels about dealing with the counterfactual situations that Fodor envisages, with their "potential" content and "hypothetical" truth conditions, i.e., what the content of certain mental states *would have been* if they had been in the

same context (which will involve intuitions regarding the importance of causal history to the content of intentional states).

Therefore, in this chapter Fodor's indexical mapping theory of narrow content has been explicated by following through his reasoning about the issues involved. In the next chapter various elements of the mapping theory are assessed and criticised.

## Notes to Chapter 4

<sup>1</sup> Over the years Fodor has had various dalliances with the notion of some kind of narrow content in relation to the standard Twin Earth thought-experiments. In Fodor (1982) he proposed that the twins' water/Twin-water beliefs would share a "phenomenological belief", wherein *x* is water if *x* is clear, transparent, drinkable, fills the oceans and rivers, etc. This idea also echoed his 1987 view in that the evaluation of the phenomenological belief was to be context-sensitive, viz., a "Principle of Reasonableness" was to function in order to produce the appropriate "universe of discourse" in which to evaluate the phenomenological belief (1982, p.111). It seems to me that this view also resembles a narrow representational view of content, which will be examined in Part 3. In a somewhat strange move, Fodor (1994) argues that a notion of narrow content may be superfluous to the needs of an explanatory psychology, so that a wide content psychology may be adequate. He comes to this conclusion by arguing that the two usual reasons for positing a narrow content, i.e., Twin cases and Frege cases, may be able to be explained away. He does this by basically arguing that standard Twin cases are nomologically impossible and that Frege cases are just rare occurrences which do not generally affect the effectiveness of a wide content psychology. However, this seems a dubious strategy for Fodor to adopt, and it has received criticism; for example, see Cheng (2002) in relation to his claims about Twin cases and Wakefield (2002) in relation to his claims about Frege cases.

<sup>2</sup> See Burge (1986) for a good discussion of whether the behaviour of subjects in Twin Earth thought experiments should be counted as identical.

<sup>3</sup> This is from Macdonald, 1995, pp.164-165.

<sup>4</sup> Fodor (1991a) again employs this "cross-context" test for determining the identity of causal powers. However, here he does admit that something like the externalist viewpoint just mentioned in the main text may prove that the causal powers of me and my twin's "Bring water" are not the same, even in the same context. The upshot of this is that Fodor then embarks on a very complicated strategy involving distinctions between the properties of causes and which of these distinctions actually affects causal powers. There is a good discussion of the issues involved here in Macdonald (1995); and criticisms of Fodor's (1991a) strategy from, e.g., Adams (1993), Burge (1995). Also, see Barrett (1997) who argues in favour of a modified version of Fodor's cross-context test. An argument similar to the externalist one given here is raised later in relation to Fodor's (1987) notion of narrow content, the basis of which, like the above externalist argument, comes from the supposed consequences of the causal-historical theory of reference. Possible strategies for deflecting this argument will be examined then.

<sup>5</sup> Or as Burge (1986) makes the similar point: "Since propositional attitudes are among the determinants of our 'behavior'...and since propositional attitudes do not supervene on our brain states, not all determinants of our 'behavior' supervene on our brain states" (1986, p.13).

<sup>6</sup> This influence stemmed from the work of Frege (1892), and later Carnap (1947). Putnam (1975) briefly discusses the historical development of the terms "intension" and "extension" on pp.216-219.

<sup>7</sup> An indexical term such as "I", "you", etc. is usually defined as being a term whose meaning is context-sensitive, that is, when the term is uttered in different contexts, it may well have different meanings. To explain the relationship of Kaplan's notion of "character" and indexical terms it is perhaps best to consider an example. Imagine that the indexical sentence type "I am tired" is uttered by myself and another person, say, Mr X. Now, it could be claimed that the propositional content

expressed by our respective sentence utterances are different, i.e., when I utter the sentence, the proposition expressed is that “MD is tired”, whilst when Mr X utters the sentence, the proposition expressed is that “Mr X is tired”. However, if it is claimed that the meaning of a sentence is the proposition it expresses, then it seems that despite the indexical sentence “I am tired” which Mr X and I utter, being of the same type, tokens of that sentence type uttered by different people will mean different things. But as Dennett (1982) remarks “...there does seem [a]...need for a sense of ‘meaning’ according to which we can say that all tokens of a sentence type *mean the same thing*” (1982, p.16). (The above example has been influenced by Dennett’s (1982) discussion of indexicals on p.16).

This is where Kaplan (1989) comes in, where he claims that indexical sentence types like the above do have a meaning which is the same, a meaning which he calls “character”. The character of an indexical expression “...is set by linguistic conventions and, in turn, determines the content of the expression in every context”(1989, p.505); or put another way, the character of an expression can be represented as a “...function...from possible contexts to contents” (1989, p.505).

What I think this means is the following. Given the indexical sentence type “I am tired”, the character of the indexical expression “I” is that it means, or refers to, the person who has uttered or written that particular occurrence of “I”. So, if I utter the sentence “I am tired”, given the character of “I” and the context in which it is uttered (i.e., by me), the content of the indexical expression “I” will be MD. However, if Mr X utters the sentence “I am tired”, given the character of “I” and the context in which it is uttered (i.e., by Mr X), the content of the indexical expression “I” will be Mr X. Therefore, as before, when Mr X and I utter the sentence, two different propositions will be expressed, which will mean that the two sentence tokens have different meanings. However, what can now be claimed is that the two tokens of the indexical sentence type “I am tired” will still have a shared “character”, which will be a function which takes the context in which either me or Mr X is the utterer into the proposition that is true if and only if the utterer is tired at the time of utterance. (In giving a specification of Kaplan’s character, I have been greatly helped by Stalnaker, 1989, pp.175-176 and Block, 1991, p.34).

Essentially, what Fodor does is to extend Kaplan’s notion of character to apply to thought-contents. The content of a thought is a function from the context of the thought onto the truth conditions the thought would have in that context. So, whereas Kaplan’s character of an indexical word, such as “I”, is determined according to linguistic conventions, it seems that Fodor’s definition of content is dependent on whatever environmental factor is relevant to the specification of the content that the thought would have in a particular context (from Stalnaker, 1989, p.176).

<sup>8</sup> As previously mentioned, White’s (1982) conception of narrow content or what he calls “partial character”, is also based on Kaplan’s character. It seems to me that the most pertinent difference between Fodor’s conception and White’s conception is that the latter involves a context of acquisition. With Fodor, what he had in mind is that the narrow content of a thought is a function which maps the thought and the context of thought onto truth conditions. With White, the narrow content or partial character of a thought is a function which maps the thought and the context of acquisition for the words expressing the thought onto truth conditions. So, whereas with Fodor’s conception, it is just imagined that either I am having my thought in the Twin Earth-context of my twin, or my twin is having his thought in my Earth-context, with White, he imagines that I grow up and acquire the relevant word in the Twin Earth-context of my twin, and vice versa for him in my Earth-context. Therefore, White’s partial character of “water” would be a function that, if I have grown up and acquired the word “water” on Twin Earth, would map my “water” onto XYZ, whereas, if my twin has grown up and acquired the word “water” on Earth, would map his “water” onto H<sub>2</sub>O. This means that White could still have Fodor’s extensional identity criterion for narrow content, but the contexts involved would be ones of acquisition rather than utterance or thought, i.e., two thoughts have the same narrow content if they effect the same mapping of thoughts and contexts of acquisition onto truth conditions. The ramifications of whether the context involved with Fodor’s mapping theory of narrow content is one of thought or acquisition will come to the fore when Block’s (1991) objection to Fodor’s theory is considered later.

<sup>9</sup> A point worth mentioning is that Fodor’s mapping theory of narrow content which we have been examining (and I suppose, narrow content construals based on the notion of Kaplan’s character in general), only seem to concern Twin Earth cases. By this, I mean thought experiments where there is a difference in the wide content or truth conditions of a pair of twins’ neurophysically identical beliefs, in their respective contexts. Specifically, there has been no application of the mapping theory to so-called Frege cases; which roughly involve a situation where a subject believes Fa but does not believe Fb, even though a = b. It will be recalled that in the previous part on conceptual role construals of narrow content, we were positively knee-deep in Frege cases, particularly in relation to Loar’s notions of psychological and social content. There is a good reason for this, which is basically that Fodor’s



mapping theory won't work for Frege cases. Consider A and B in the same context, A has the belief that the Morning Star is bright whilst B has the belief that the Evening Star is bright. The mapping theory states that two beliefs have the same narrow content if they effect the same mapping of beliefs and contexts onto truth conditions. This means that because A's belief and B's belief in the same context are both mapped onto the same truth condition, i.e., that Venus is bright, then according to the mapping theory, both beliefs have the same narrow content, which doesn't seem quite right. This point is raised, and examined in more detail, by Aydede (1997), who goes on to observe that even though Fodor (1994) discusses Frege cases while rejecting narrow content as superfluous to psychological explanation, the notion of narrow content that he rejects is not the mapping theory version that he has been mainly concerned with in the years 1985-1991 (1997, p.434).

Aydede's point about Fodor's mapping theory and Frege cases seems similar to the point that Stich (1991) makes concerning Fodor's narrow content and so-called fat syntax. Stich's objection put crudely, is that the mapping theory narrow content will be too coarse-grained to capture cognitive generalisations which are described in a syntactic taxonomy of mental states (Stich, 1991, p.249). Therefore, as Fodor favours a computational theory of mind (CTM) which views mental states and processes in syntactical terms, his notion of narrow content will not be capable of stating "...generalizations that describe the mind's workings..." (Stich, 1991, p.249).

<sup>10</sup> Perhaps this situation could also be described in the following manner: the same thoughts could have the same truth conditions, but the same thoughts may not have the same truth values.

<sup>11</sup> However, it still seems plausible that the narrow content could be expressed as a Russellian definite description, e.g., the x such that x is transparent, colourless, odourless, liquid...etc.

<sup>12</sup> It is worth mentioning here that Fodor's conception of narrow content is also amenable to his "language of thought" hypothesis, which again may well have motivated his view of narrow content (this point is mentioned in Stalnaker, 1989, p.180).

## CHAPTER 5

### ASSESSMENT AND CRITICISMS OF FODOR'S NOTION OF NARROW CONTENT

Having set out Fodor's (1987) indexical conception of narrow content, it is now time to examine some of the criticisms of that conception.

#### 5.1 What is a Function? – (i)

Stalnaker (1989) questions how Fodor intends to specify the mapping of thoughts and what exactly is the “function” which is involved in the same mapping of thoughts and contexts onto truth conditions, e.g., “...how it is that the relevant function is determined by what is in the believer's head” (Stalnaker, 1989, p.176). He then gives a “parody” of Fodor's extensional identity criterion for narrow content, the gist of which is the following. Stalnaker imagines a world where he is located exactly three miles from a burning barn, whilst on a counterfactual world, his counterpart is in the same position, but is located exactly three miles from a snow-covered chicken coop (Stalnaker, 1989, p.177). He then claims that there is presumably something about the relationship between himself and his world which entails that he is three miles from a burning barn, which Stalnaker calls condition C'. Also, there is something about the relationship between his counterpart and his world which entails that he is

three miles from a snow-covered chicken coop, which Stalnaker calls condition C

(Stalnaker, 1989, p.177). He then declares that:

Whatever these conditions are, we *do* know this: short of a miracle, it must be true that anyone in the location that both I and my counterpart are in our respective worlds would be three miles from a snow-covered chicken coop if condition C obtained, and three miles from a burning barn if instead C' obtained. But this does not help us identify a specific function that takes condition C' into the property of being three miles from a snow-covered chicken coop and also takes C into the property of being three miles from a burning barn – a function that is supposed to represent the contribution that an individual's location makes to the relational property. There are many such functions, and no reason to identify any of them with the contribution that my intrinsic location makes to the specific relational property. My counterpart cannot reasonably say, "I did my part toward being three miles from a burning barn by going to a place where, if condition C' had obtained instead of C, I would have been three miles from a burning barn." *Every* location is such that for some external conditions, if those conditions obtain, then anything in that location is three miles from a burning barn (Stalnaker, 1989, p.177).

It will be recalled that Fodor's narrow content criterion which is being parodied here, is one where there is claimed to be something about the relationship between my twin and Twin Earth which resulted in his "water"-thoughts being about XYZ; and something about the relationship between me and Earth which resulted in my "water"-thoughts being about H<sub>2</sub>O. The former relationship of my twin is called condition C, whilst my latter relationship is called condition C'. Fodor then claimed that "short of a miracle", if I satisfied condition C along with my twin, or if my twin satisfied condition C' along with me, in both situations, me and my twin's water thoughts would share their truth conditions (or would have the same extensions or wide content). This then led to Fodor's extensional identity criterion for narrow content, i.e., two thoughts would have identical narrow contents if they effect the same mapping of thoughts and contexts onto truth conditions.

Of course, Stalnaker's parody is just that, a parody, so it is not meant to be exactly analogous to Fodor's narrow content criterion. As always with Stalnaker, his points are important and well made, but I think that the disanalogies between his parody and Fodor's narrow content criterion may give an answer to his criticism. It seems to me that the main point that Stalnaker is trying to make is that just due to their intrinsic locations, there is no specific function that two subjects could share, that would take them into being either three miles from a snow-covered chicken coop or three miles from a burning barn, even if they both satisfied either condition C or C'. That is, if two subjects satisfied condition C, then they would both be three miles from a snow-covered chicken coop, and if two subjects satisfied condition C', then they would both be three miles from a burning barn. However, what narrow function could the two subjects be claimed to share because of the above claims; a function that maps a subject's location onto a snow-covered chicken coop in one situation, and maps a subject's location onto a burning barn in another situation. Therefore, how can Fodor claim that there is a specific function that the "water"-thoughts of two subjects share, which will result in their thoughts being mapped onto either H<sub>2</sub>O or XYZ in different situations, just because both subjects' thoughts would have the same truth conditions if they both satisfied either condition C' or C?

A major disanalogy is that Stalnaker's parody does not involve any intentional states of the subjects at all, only the physical locations of the subjects. Commenting on Stalnaker's parody, White (1992) states that:

*Believing* that one is in the vicinity of water (and not XYZ) is not like *being* in the vicinity of water. Absolutely anything, regardless of its intrinsic nature will be in the vicinity of water if it is in the right external circumstances (namely, if there is water in its vicinity). But it is not the case that anything,

regardless of its internal makeup, will believe that it is in the vicinity of water merely in virtue of having the right relations to the environment – at least where these relations do not entail that it has an appropriate internal structure (White, 1992, p.53).

In claiming this, White is answering what he takes to be an “uncharitable interpretation” of Stalnaker’s parody, which involves “...skepticism about the assumption that the narrow properties of a subject make a necessary contribution to the content of the subject’s beliefs” (White, 1992, p.53). I think that White’s point is not only that, say, a boulder, could be in the vicinity of water due to its physical location, but not believe that it is in the vicinity of water because it does not have the appropriate internal structure. A human being who has the appropriate internal structure could be in the vicinity of water and yet not believe that he was in the vicinity of water, e.g., because he did not know it was there. White’s point is that someone would also need the appropriate internal functional structure to have the belief that he was in the vicinity of water (White, 1992, p.53). Whilst it is fair to say that Fodor would probably not agree with White on this point, it could still be in the right direction towards answering Stalnaker’s criticism.

That is, what picks out the specific function are the internal narrow properties of the subject, such as conceptual and perceptual properties. For instance, one could imagine that a subject’s belief about water would act as a function which would map onto H<sub>2</sub>O on Earth and XYZ on Twin Earth. The function of water would map onto the transparent, odourless, potable liquid that fills the oceans and lakes, comes out of taps, etc. in each context. Of course, the point to note here is that even though the wide content of “water” is different on Earth and Twin Earth, the conceptual and perceptual properties of “water” in each context are the same, which must play some

part in the mapping of the “water” thought-token onto its extension or truth conditions in the respective contexts.

This is important with regards to Stalnaker’s parody of Fodor’s position. In the parody, even if Stalnaker had the *belief* that he was three miles from a burning barn and his counterpart had the *belief* that he was three miles from a snow-covered chicken coop, these beliefs would not provide a narrow shareable function that would map onto the burning barn in one context and the snow-covered chicken coop in another context. As with the previous water-beliefs, the wide semantic content of each belief is different in each context, but the conceptual and perceptual properties of the beliefs are also different in each context, which means that there will not be a narrow function shared by the beliefs which will map the belief onto a burning barn in one context and a snow-covered chicken coop in another context.

Therefore, to answer Stalnaker’s question how the particular narrow function is picked out and specified, one can say the following. The function of a subject’s water belief will depend not only on the wide content of water on Earth or the actual world, but also on the conceptual and perceptual properties associated with water in that context. It seems plausible that these conceptual and perceptual “water” properties then play some role in mapping the water belief onto perhaps a different wide content in another context, but a wide content that has the same conceptual and perceptual properties as the appropriate wide content on Earth or in the actual world. Whether these conceptual and perceptual properties play a constitutive or individuating role with the relevant beliefs, and what this means for the content of those beliefs, will be examined in the next section that deals with the objections of Block (1991).

## 5.2 Does Fodor's Narrow Content Collapse Into Syntax?

By far the most detailed and involved critical examination of Fodor's mapping theory<sup>1</sup> of narrow content has been by Block (1991) and it is this work which I shall now examine. What I take to be Block's main objection to Fodor's mapping theory is that it collapses into syntax. However, before explaining and addressing this objection, I will examine Block's discussion of what he calls the "mapping theory" that Fodor is utilising in his narrow content conception, as this may have some bearing on the resultant objection that Block makes. He claims that:

The most straightforward application of the model of character to narrow content is that the narrow content of the thought expressed by my utterance of "Water puts out fires" is the mapping from contexts (actual and possible) of acquisition of my words to the contents (truth conditions) of the thoughts that would be expressed by the utterance of those words in those contexts (Block, 1991, p.34).

Block calls the above the "most straightforward application", but it is interesting to note that this may not be the application that Fodor intended with his conception. From what has gone earlier, Fodor's conception may be summarised as that the narrow content of two thoughts is the function that effects the same mapping of thoughts and contexts onto truth conditions. That is, if my thought that "water is wet" occurs in the context of Twin Earth, my thought would be mapped onto the truth condition or content that "XYZ is wet" and if my twin's thought that "water is wet" occurs in the context of Earth, his thought would be mapped onto the truth condition or content that "H<sub>2</sub>O is wet". This conception does admittedly have its vague points, but it is a questionable point as to whether the "contexts" involved in this also involve contexts of acquisition. *By "context of acquisition", Block seems to mean a situation*

*where one is brought up in the context concerned and physical/socio-linguistic factors contribute towards one's learning what a particular word means in that context* (Block, 1991, p.37). That is, a context of acquisition is a context of “linguistic” acquisition and also a context of “acquaintance”, where the person concerned has causal interactions with the kind in question.<sup>2</sup> Indeed, Block himself does acknowledge that Fodor does not actually claim that his conception of a “context” does involve a context of acquisition (Block, 1991, p.35). Moreover, Fodor (1991b) himself, in a reply to Block's paper, claims that he doesn't “...hold a “context of acquisition” account of broad content...I [Fodor] don't think that being acquired can be an essential condition for having content” (Fodor, 1991b, p.n314).

Now, I admit that Block is quite correct to claim that the way Fodor (1987) talks of words such as “water” being anchored to the content they have in a particular context does give support to his claim that Fodor's conception involves a context of acquisition (Block, 1991, p.35). Block gives an example of an Earthian crash-landing on Twin Earth and not being able to make true claims or statements about the “water” he finds there, because the water on Twin Earth is a different stuff to the water on Earth (Block, 1991, p.35). I think that Fodor would agree with Block on this point, and with the point generally in relation to Earth/Twin Earth “switching” or “visiting” cases in general. However, I would claim that it still does not necessarily follow that Fodor takes the view that his conception of a context in relation to the narrow content mapping theory also includes a context of acquisition.

The reader may be wondering why I am belabouring this point about a context of acquisition for words and expressions, what difference does it make to Fodor's



mapping theory conception of narrow content? The reason is that when the context of acquisition is included as part of the “context” of the mapping theory, it is this that Block uses to produce his objection that narrow content collapses into syntax. This occurs roughly as follows. Imagine the thought which I express by the utterance “Mud makes a good shampoo” in the context of acquisition of the actual world. So far so good. But now Block claims that what my twin means by uttering “Mud makes a good shampoo” “...in various contexts of acquisition will depend *entirely on[his] upbringing in the physical and social world of the contexts* in which [he] acquire[s] language” (Block, 1991, p.37).

For example, in the context of acquisition of another possible world, the utterance of “Mud makes a good shampoo” might mean Eskimos never eat the yellow snow. What this means is that the thought “Mud makes a good shampoo” would, in the different possible context, be mapped onto a wide truth conditional content that is radically different to the wide truth conditional content that the thought has in the context of the actual world. The upshot of this is that the narrow content of the thoughts “Mud makes a good a shampoo” would only have an identity of syntax, i.e., both thoughts would only have the same syntactical shape and structure of their contents in common.<sup>3</sup> This effectively means that the narrow content in question is too coarse-grained to provide an adequate shared psychological explanation of me and my twin’s behaviour of washing our hair in mud, as the wide content of my twin’s utterance would seem to have no relevance to his consequent hair-washing behaviour (Block, 1991, p.37).

This then is roughly the method by which Block constructs what I take to be his main objection against Fodor's mapping theory conception of narrow content, i.e., the objection that the narrow content collapses into syntax. It is now time to consider some responses that can be made to Block's objection. Firstly, if we do accept Block's claim that a context of acquisition is to be included in the "context" of the mapping theory, then the most obvious response seems to me to be one which basically stipulates that the subject's conceptual and experiential elements connected to the subject's particular expression/thought in question are held constant from the actual context of acquisition to the counterfactual context of acquisition for that particular expression/thought. An important point should be emphasised here. With the above, I am not stipulating that the actual context of acquisition of an expression/thought is the same as the counterfactual context of acquisition of the expression/thought. What would be stipulated is that the twins' conceptual and experiential elements which are connected to the syntactic shapes of the actual and counterfactual expressions/thoughts are kept constant.

However, despite this, the actual and counterfactual contexts of acquisition could still be different as they would have different wide truth conditional content. For example, with my thought that mud makes a good shampoo in the actual context and my twin's thought that mud makes a good shampoo in the counterfactual context, from the subjective first-person perspective of how me and my twin conceive things, the syntactic shapes of the thoughts would have the same conceptual and experiential factors connected to them, i.e., we would have the same "superficial" concept of "mud" as being a mixture of water and earth, of "shampoo" as being a substance one washes one's hair with, etc. However, if water in the actual context of acquisition is

composed of H<sub>2</sub>O and water in the counterfactual context of acquisition is composed of XYZ, then our contexts of acquisition would still be different for our respective “water” terms. The result would be that my thought that mud makes a good shampoo would be true if and only if mud makes a good shampoo, whilst my twin’s thought that mud makes a good shampoo would be true if and only if, say, Twin-mud makes a good shampoo. Therefore, the subjective conceptual and experiential elements associated with the thought in the actual world are carried over to the thought in the counterfactual world, even though the “meaning” of the thoughts, in a wide truth conditional sense, are different, due to physical differences in the environments of the respective contexts of acquisition of the thoughts.

This does not seem an unfairly restrictive stipulation to me, but rather it just involves the normal situation that is usually utilised in relation to thought experiments in the philosophy of mind. For example, with Putnam’s (1975) original thought experiment, the subjective conceptual and experiential elements for Twin-Oscar’s “water” in the context of acquisition of Twin Earth are just stipulated as being the same as Oscar’s “water” in the context of acquisition of Earth. That is, the narrow meaning of the respective “water” terms have the same conceptual and cognitive values to the doppelgangers, it is only the wide truth conditional content of the water terms which are different, with water on Earth being composed of H<sub>2</sub>O and water on Twin Earth being composed of XYZ. With Burge’s (1979) arthritis thought experiment, he also explicitly stipulates that the sensory and physical factors related to the actual and counterfactual “arthritis” thoughts are the same (Burge, 1979, pp.77-78). This is despite the fact that the actual and counterfactual contexts of acquisition for “arthritis” are different, this time due to differences in the social environments of the respective

contexts. Namely, in the actual context “arthritis” means a rheumatoid inflammation of the joints only, whilst in the counterfactual context “arthritis” means a rheumatoid inflammation of the joints and bones. However, the point is that Burge gets to freely stipulate that the cognitive and experiential elements connected to the “arthritis” thoughts in their respective contexts of acquisition are going to be the same in his thought experiment.

What Block is suggesting in relation to Fodor’s mapping theory conception seems to me to be fairly equivalent to someone saying to Putnam: “Sorry, but on Twin Earth there is a radically different context of acquisition for “water” than there is on Earth. On Twin Earth “water” means my dog has no nose, so your thought experiment isn’t as plausible as it seemed in relation to the claim about differences in the physical environment affecting the thought-contents of doppelgangers”. One could also imagine someone saying to Burge: “Sorry, but in the counterfactual situation of your thought experiment, the context of acquisition for the expression “arthritis” is the same as it is in the actual situation, so that counterfactually “arthritis” also means an inflammation of the joints only, so there is no change in the actual and counterfactual contents of arthritis thoughts”.

In situations like these, I think it is fair to say that Putnam and Burge would be quite justified to tell the somebody concerned to stop questioning their respective stipulations for the contexts of acquisition in their thought experiments. I feel that is pretty much the same situation with Fodor and the contexts of acquisition in his thought experiment. Fodor is trying to specify what the narrow content would be that is shared between the thoughts of intrinsic physical duplicates. In order to do this he

stipulates that the subjective conceptual and experiential properties of the relevant actual and counterfactual expressions/thoughts are the same, whilst then stipulating that the actual and counterfactual contexts of acquisition for the expressions/thoughts are different in a specified way, i.e., different wide truth conditional content for those expressions/thoughts.

### 5.2.1 *What is a Function? – (ii)*

However, what if one does not just stipulate that the subjective conceptual and experiential properties of the thoughts in question remain constant, in answer to Block's objection of the collapse of the mapping theory's narrow content into syntax. In this case, the most obvious response is to examine what exactly is meant by the "function" which is involved in the mapping of thoughts and contexts onto truth conditions. Recall that Fodor claims that the narrow content of two thoughts is that function which maps the thoughts and contexts onto the same truth conditions. Another way of putting this is supplied by Loewer/Rey (1991) where, in relation to Fodor's conception of narrow content, they state:

...a context is any circumstance in which a locking pattern constitutive of wide content obtains. Narrow content is that state of an organism that brings about such locking. Thus...the earth is such a context in which 'water' is locked onto water, and Twin-earth is one where it is locked onto XYZ. What Sophie and her twin share is a state disposing their symbols to become so locked (Loewer/Rey, 1991, p.xxix).

Therefore, it seems that what is needed is a specification of the narrow function that results in the relevant mapping or locking of a pair of thoughts and contexts onto the

same truth conditions. This could be achieved by putting a constraint on the content of the belief that is realized in a possible world, so that its content could not be something outlandish. As suggested in section 5.1, perhaps this could involve the conceptual and experiential elements involved in having the belief, which helps individuate the agent's belief from the subjective first-person perspective of the agent himself. These conceptual and experiential elements could be included either directly in the notion of the belief's realization conditions, or as an extra component added to the individuation conditions of the belief, perhaps by using a variation on Edwards' (1994) idea of an "experiential component" or "E-component".

Therefore, let us consider where we are at this point. There is the Earth-context, where, say, my "mud"-thought is mapped onto mud in that context. There is also the Twin-Earth context, where Twin-me's "mud"-thought is mapped onto Twin-mud in that context. According to Fodor, the narrow content of the pair of thoughts is the function that will map the thoughts and contexts onto the same truth conditions or wide content. This could have the schema (following Edwards (1994)) that:

(a)  $C(\text{thoughts, context}) > \text{truth conditions}$

so that with the contexts of Earth and Twin Earth we have

(a1)  $C(\text{mud makes a good shampoo, Earth}) > \text{mud makes a good shampoo}$

(a2)  $C(\text{mud makes a good shampoo, Twin Earth}) > \text{Twin-mud makes a good shampoo}$

Therefore, if I am on Twin-Earth my “mud”-thought will be mapped onto the same truth conditions as the “mud”-thought of Twin-me in that context. Or, if Twin-me is on Earth his “mud”-thought will be mapped onto the same truth conditions as my “mud”-thought in that context. It can be seen that the truth conditions of our respective “mud”-thoughts in the Earth and Twin Earth contexts are different, this is due to Earth mud being composed of water and earth and Twin Earth mud being composed of XYZ and earth. Block’s objection to Fodor’s claim, as I understand it, is the following. He is claiming that the Twin Earth context of Twin-me should include a context of acquisition for the words and meanings that are used in that linguistic community. If this is accepted, then he further claims that the context of acquisition for Twin-me’s thought that mud makes a good shampoo could be something completely different to what it is for my thought on Earth. That is, even though our mud thoughts would be syntactically identical, Twin-me’s mud-thought could be mapped onto different truth conditions or wide content, e.g., that grass is green, or something like that. The upshot of this would be that if I had my “mud”-thought in the Twin Earth context, it would not be mapped onto the same truth conditions/wide content as Twin-me’s “mud”-thought in that context, and vice versa for Twin-me’s mud-thought in the Earth context. This enables Block to claim that mapping theory narrow content is too coarse-grained to be acceptable for a psychological explanation of me and my twin’s subsequent behaviour.

It seems to me that to counter Block’s present objection, an inclusion of conceptual and experiential properties are going to be required in the specification of the function involved in the mapping of the relevant thoughts and contexts onto truth conditions.<sup>4</sup>

Perhaps one could say that the mechanism of a “function” is similar to that of a Fregean “sense”<sup>5</sup> or a Carnapian “intension”, in that both these entities supposedly determine the extension of a thought.<sup>6</sup> However, to illustrate this notion, let us first examine the normal picture of how a function maps a thought onto a truth condition, e.g., consider me having the thought which I express with the utterance that “Mud makes a good shampoo”. On a truth conditional theory of meaning and content, the content of my thought is considered to be a function from thoughts to truth conditions, which can be illustrated (again following Edwards (1994)) as :

(b) C(thoughts) > truth conditions

where the content is going to be wide, as the truth conditional element of the function is dependent on objects and states of affairs in my actual environment. Therefore, if we put my particular thought into schema (b), we have:

(b1) C(Mud makes a good shampoo) > Mud makes a good shampoo

So the content of my thought “Mud makes a good shampoo” is a function from that thought to the truth conditions for that thought in the actual world, which is that mud makes a good shampoo. This isn’t a very exciting result, but it does enable one to consider what the mapping mechanism of the “function” involved here consists of. In doing this, I will be presupposing something along the lines of Fodor’s (1975) Language of Thought (LOT) hypothesis, which basically views beliefs and other propositional attitudes as particular relations to mental sentences in the brain.<sup>7</sup> For instance, according to the LOT, my believing that mud makes a good shampoo would



mean that I had a “belief” relation to the mental sentence “mud makes a good shampoo”. Moreover, the mental sentences (and I suppose the mental words which make them up) in the LOT are considered by Fodor to be representational, that is, they represent objects and events in the external world.

Indeed, it seems plausible to me to consider the mental sentence “mud makes a good shampoo” as having a total intension, which is a composite of the intensions of its subsentential parts.<sup>8</sup> That is, the total intension for “mud makes a good shampoo” will be composed from the intension of “mud”, the intension of “shampoo”, etc. Again it seems plausible to suppose that, say, the intension of “mud” will involve conceptual properties, such as mud being a mixture of earth and water, and the various conceptual relations to other thoughts that this mud intension will have. In addition to this, the intension of “mud” could also involve experiential properties, i.e., my sensory perceptions of mud, my mental images of mud, etc. Finally, this conceptual/experiential intension for my Mentalese “mud”, could be said to determine its extension in the world, or rather, to map my “mud” thought-token onto the mud-stuff in my actual environment. By this method, one could imagine the thought “mud makes a good shampoo” as having a composite total intension which would map the thought onto its truth conditions or wide content in my actual environment.

Of course, it should be noted here that to stay in line with Fodor’s reasoning, we should have to acknowledge that the intension of my thought in the actual world is derivative on wide content, that is, the intension has been formed using properties that are outside the head, so to speak. It seems to me that this is the main difference between Fodor’s mapping theory conception and the narrow content conception of

Loar that was looked at earlier. It will be recalled that Loar's realization conditions were considered not to have subjunctive or counterfactual content, as one did not know whether the thoughts involved were occurring in the actual or counterfactual situations. Whichever world the thought occurred in was considered to be the actual world and its content was evaluated as such, rather than fixing the content of the thought in the actual world and then evaluating what its subjunctive content would be in a counterfactual world.

This point comes to the fore when we consider Twin-me's thought that "mud makes a good shampoo" on Twin-Earth, which is physically identical to Earth except that the stuff called "water" on Twin-Earth is composed of XYZ. If we now put Twin-me's thought into schema (b) to find its truth conditional content, we get the following:

(b2)  $C(\text{Mud makes a good shampoo}) > \text{Twin-mud makes a good shampoo}$

Even though Twin-me's mud-thought has the same syntactic shape as my mud-thought, the former's mud-thought is mapped onto a different mud-stuff than my mud-thought is, so the truth conditions of Twin-me's mud-thought is different to mine. This situation has come about due to the content of my "mud" being fixed by whatever it is in the actual world (earth and water), and so the content of Twin-me's "mud" will be evaluated as subjunctive content (earth and Twin-water).

Of course, it is at this point, when considering the mud thought of my twin, that Block's objection of narrow content collapsing into syntax is raised. He argues that the context of acquisition of my twin is different to mine and his thought "mud makes

a good shampoo” means something completely different in that context. The question now is whether my intensional construal of the function involved in the mapping of thoughts onto truth conditions is constraining enough to deflect Block’s objection.

Personally, I think that it is, in the following way. With the total intension of my thought that mud makes a good shampoo on Earth involving the conceptual and experiential factors involved in having the thought, my hope is that these factors will act as additional individuation conditions for the thought in question. On the Earth, the intension of my mud-thought would map my “mud” token onto mud (i.e., earth and water) in that context. However, further to this, my concept of mud would be that it is a mixture of earth and water, that it has a particular perceptual appearance, that using mud as a shampoo would involve washing my hair with mud and so on. What my mud-thought meant to me would involve all the above factors. Of course, most of these factors I acknowledge are derivative on wide content, and plausibly enough on the context of acquisition for my use of language on the Earth. However, the important point is that given my thought that mud makes a good shampoo in the context of the Earth, then for me, that thought has the particular meaning that it does for me, at that moment. When it is then imagined that at that exact moment on Twin Earth, Twin-me is also thinking that mud makes a good shampoo, this thought is not just syntactically identical to mine, but the thought’s intension, from the subjective perspective of Twin-me, is also identical to mine.

What I mean by the latter claim of intensional identity is the following. Twin-me’s thought that mud makes a good shampoo will have a total intension that results in Twin-me believing that mud is a mixture of earth and water, that it has the same

perceptual appearance that mud has to me, that using mud as a shampoo would involve washing his hair with mud, etc. Now, to this claim one could imagine the objection being made that the intension of Twin-me's mud-thought is not the same as mine, due to the fact that what Twin-me believes is that mud is a mixture of Twin-water and earth, which results in him having a Twin-mud thought rather than my mud-thought.

However, the point to make is that from the subjective first-person perspective of me and Twin-me, our respective intensions for our "mud" thought tokens are each mapped onto the local mud-stuff in our respective environments. Moreover, both me and Twin-me's concept of mud as being a mixture of earth and water will also be intensionally identical from the subjective perspective, as our "water" thought tokens will map onto the local watery-stuff in our environments. Of course, the extensions of our respective "mud" thought tokens will be different on the Earth and Twin Earth, but this does not entail that the subjective "mud" intensions that pick out those extensions must also be different. Indeed, with so-called intensional semantics, it is acknowledged that the same intension in a different possible world may pick out a different extension to the one it picks out in the actual (or another possible) world.

Therefore, with regards to the subjective first-person "mud" and "water" intensions of me and Twin-me, one could say that what is required is that the definite descriptions or stereotypes that make up these intensions are fixed between Earth and Twin Earth, whilst the chemical essences of the relevant stuffs picked out by the intensions may change. If all this is accepted, the result will be that the intensions of me and Twin-me's "mud" thoughts will share the same conceptually and experientially fixed

properties; which in turn will mean that Twin-me's mud-thought will have the same fixed subjective first-personal meaning that my mud-thought has to me, i.e., a meaning based on our fixed (and so shared) conceptual and cognitive properties. This hopefully means that Twin-me's "mud" thought-token will be *validly* defined on Twin Earth, i.e., it will be mapped onto Twin-mud, just as my "mud" thought-token is mapped onto mud on the Earth.

If there is a situation like Block envisages, where Twin Earth has a different context of acquisition for the term "mud", which results in Twin-me's "mud" thought-token being mapped onto an outlandish truth condition or wide content, then this would be considered an *invalidly* defined mapping. The latter would mean that Twin-me's "mud" thought-token is not mapped onto a wide content that has the same conceptual and experiential properties as the wide content that my "mud" thought-token has on Earth. To put it another way, Twin-me's "mud" thought-token would be mapped onto satisfaction conditions, as well as truth conditions.<sup>9</sup> If the wide content that Twin-me's "mud" thought-token is mapped onto satisfies the conditions of having the same conceptual and descriptive stereotype as the wide content that my "mud" thought-token is mapped onto, then this will be a validly defined mapping. However, if the wide content that Twin-me's "mud" thought-token is mapped onto does not satisfy the conditions of having the same conceptual and descriptive stereotype as the wide content that my "mud" thought-token is mapped onto, then this would be an invalidly defined mapping. Finally, if Twin-me's "mud" thought-token is not mapped onto any wide content at all, one could say that the mapping was undefined in this context.

By this method, Block's objection that the mapping theory of narrow content collapses into syntax can be deflected. What I have tried to show is that the intensional mechanism of the "function" involved in the mapping of thoughts onto truth conditions on Twin Earth can be fixed (in relation to its conceptual and experiential properties) to what it is on the Earth. This means that the conceptual and experiential stereotypes involved in the intension of Twin-me's mud-thought must be the same as my mud-thought on the Earth, otherwise the mapping involved with the former's thought will be invalid. This then is the constraining element of the function that stops Twin-me's mud-thought from being mapped onto an outlandish wide meaning and content. The upshot of this will be that if I am in the Twin Earth context, the intension of my mud-thought will map it onto Twin-mud, just as Twin-me's intension will map his mud-thought onto Twin-mud in this context. Similarly, if Twin-me is in the Earth-context, the intension of his mud-thought will map it onto mud, just as my intension will map my mud-thought onto mud in this context.

Therefore, contra Block, when me and Twin-me are in the same contexts, our intensional functions that map our mud-thoughts onto wide truth conditional content will not be too coarse-grained to provide an acceptable narrow psychological explanation of our behaviour. When we are both either in the Earth or Twin Earth contexts, me and Twin-me's subsequent behaviour of washing our hair in mud can be subsumed under the narrow psychological explanation that we both believe that mud makes a good shampoo.

The outcome of the above reasoning seems to be that the "context" involved with the mapping theory of narrow content, i.e., the function which maps *contexts* and

thoughts onto truth conditions, is one that does not involve a context of acquisition for the thought, but that just involves a context of thought. When I am in Twin-me's Twin-Earth context or Twin-me is in my Earth-context, and we both have our respective thoughts, then the narrow content of our thoughts is that intensional-function which maps those thoughts and the context where those thoughts occur in, onto the same truth conditions. There does not seem to be a context of acquisition involved in this mapping, the thoughts just have the meanings they have at that moment of thought, which is then mapped onto the truth conditions.

What I am trying to say is that there is no historical element involved in the context, the thoughts and expressions just have the contents they do in the here and now, irrespective of the acquisitional process of how they came to have those contents. This view also seems to accord with Kaplan's notion of the "character" of an indexical expression, which was the inspiration for Fodor's (and White's) narrow content mapping theory. For Kaplan, the character is considered to be a function from contexts of utterance to contents (or truth conditions). For example, consider A and B both saying "I am now in the bath". The propositional contents (truth conditions) of these two utterances are different, but Kaplan claimed that the character of the two utterances is the same. That is, with the indexical "I", there is the same mapping from the contexts of utterance to contents or truth conditions. In both cases, the "I" maps the utterer and time of utterance into the proposition that is true if and only if the utterer is in the bath at the time of utterance.<sup>10</sup>

Therefore, it can be seen that the context involved with Kaplan's character is a context of utterance. Of course, as Stalnaker (1989) points out, there are big

differences between Kaplan's original idea of "character", and the variation of it that is involved in Fodor's mapping theory. For one thing, Kaplan's theory of indexicals is a linguistic theory, where the character of an indexical expression such as "I", is determined by linguistic conventions (in the case of "I" to refer to the speaker or writer of the "I" in question). With Fodor's mapping theory, he is trying to apply the notion of character to thoughts, with the character or function of a thought being composed out of the causal relations between the subject and his/her environment. This is a rather large jump from what Kaplan originally had in mind, and it perhaps doesn't include the extra conceptual/experiential factors which I added above to help constrain the individuation of the thought in question. Nevertheless, it still seems plausible to me to just have a context of thought involved with the mapping theory, rather than a context of acquisition.<sup>11</sup> Just as linguistic convention fixed the character of indexical expressions, could not one say that conceptual and experiential factors help to partially fix the character or intensional-function of a thought such as "mud makes a good shampoo"?

### **5.3 Objections to the Modified Construal of Fodor's Mapping Theory**

Therefore, I have now given my modified construal of Fodor's mapping theory of narrow content, which has the aim of hopefully deflecting what I take to be Block's main objection to this theory, i.e., that the narrow content involved collapses into syntax. However, there are a number of issues raised by Block's (1991) paper which raise further objections even to my modified construal of Fodor's mapping theory. What I take to be the two main objections concern whether the mapping theory



involves a context of acquisition as well as a context of thought/utterance, and whether conceptual and experiential factors could be used to successfully constrain the intensional-function of Twin-me's "mud"-thought on Twin Earth. I will first examine the objection involving the context of acquisition and context of thought.

### 5.3.1 *Contexts of Acquisition and the Causal-Historical Theory of Reference*

As seen above, I have argued that the context involved with the mapping theory need only include a context of thought, and that including a context of acquisition is the method by which Block is able to get his narrow-content-collapsing-into-syntax objection started. I previously argued that if there is a context of acquisition for the relevant thoughts involved in the mapping theory, then this should be considered the same as the context of acquisition for the relevant thoughts on the Earth (or in the actual world); as this is what is allowed to happen in the typical thought experiments of Putnam, for example.

I also argued that if constraints involving conceptual and descriptive stereotypes of the relevant thoughts involved with the intensional-function are fixed on the Earth, then this should also constrain the subjective conceptual and descriptive stereotypes of the intensional-function on Twin Earth. The upshot of this would be that if I am on Twin Earth, there would be a context of thought for the relevant thought, where the intensional-function of the thought would map it onto a particular wide content at the moment it is thought (the wide content mapped onto, would of course be the same as the mapping for Twin-me's thought). Similarly, if Twin-me is on the Earth, there

would be a context of thought for the relevant thought, where the intensional-function of the thought would map it onto a particular wide content at the moment it is thought (again, the wide content mapped onto, would be the same as the mapping for my thought). In both cases, there would only be a context of thought involved with the relevant thoughts in that context.

However, we now come to Block's objection to the above picture, where he claims that:

Suppose that I travel by spaceship to Twin-Earth, crash-landing in one of their "oceans". Unaware of the composition of the liquid that surrounds me, I radio home, "Surrounded by water". My message is false, however, since the stuff I am surrounded by is not water...I speak English, not twin English, and so my word 'water' is grounded in the "dominant causal source" (in Evans's terminology) of my word at home. In Harman's (1982) terminology, the "normal context" for my use of 'water' is the home context. Of course, if I stay on Twin-Earth for many years, my word 'water' will switch its reference to twater, as the dominant causal source of my word shifts...(Block, 1991, p.35).

So here, Block is claiming that when he suddenly enters the Twin-Earth context, his utterance or thought "water" will not be referring to the Twin-water he is surrounded by, but will still be referring to the water he has been causally interacting with back on Earth. Only after some period of time on Twin Earth, will his word "water" have acquired its new referent, which will then be Twin-water. The effect of this reasoning on the mapping theory is fairly clear. When I suddenly appear on Twin-Earth with my belief that mud makes a good shampoo, if there is only a context of thought involved, then my "mud" thought token will not be referring to, or be mapped onto, the Twin-mud in my environment, it will still be mapped onto the mud back on Earth. This means that my "mud"-thought on Twin Earth will not be mapped onto the same

truth conditions or wide content as Twin-me's "mud"-thought in that context, which is the basic claim of the mapping theory (and, mutatis mutandis, it is the same for Twin-me appearing on Earth and having his "mud"-thought). The upshot of this appears to be that the contexts involved with the mapping theory would have to include a context of acquisition as well as a context of thought, for me and my twin's "mud"-thoughts to eventually map onto the same truth conditions in the same context.

The first thing to say is that if Block is correct about this, and a context of acquisition has to be included in the notion of a context, along with a context of thought, then this does not spell immediate doom for the mapping theory. As stated above, when for example, I am in Twin-me's context having my "mud"-thought, it can either be stipulated that the context of acquisition for both our "mud"-thoughts are conceptually and experientially the same, or these factors can be built into the intensional-functions of the thoughts themselves. In this way, the contexts of acquisition for the "mud"-thoughts will stay the same for me and my twin, and the collapse of narrow content into syntax can still be avoided. All it would mean, according to Block's above quotation, is that after a certain period of time had elapsed, the intensional-functions of both me and my twin's "mud"-thoughts would map those thoughts onto the same truth conditions in that context.

However, even though I must admit that Block's point does seem intuitively correct, it still seems plausible to me that just a context of thought is involved, with me just suddenly appearing on Twin-Earth and my "mud"-thought being immediately mapped onto Twin-mud, just like Twin-me's "mud"-thought is. Therefore, I will make an

attempt to try to turn the intuition towards a position of supporting just a context of thought for the context involved.

To begin with, Block's quotation seems to be presupposing what is called the causal-historical theory of reference, or at least a variation of the original notion of causal-historical referring, which is usually attributed to Kripke (1980).<sup>12</sup> It is a variation, because according to Kripke's original ideas concerning the causal-historical reference of proper names and natural-kind terms, there never would be a switching of reference of the "water" term in Block's quotation, his "water"-term on Twin Earth would never refer to Twin-water. The reason for this is that according to Kripke (and later Putnam), the reference of natural-kind terms such as "water" are fixed according to an original "dubbing", where the term "water" would be first used to refer to a particular substance. All later uses of the term "water" would be connected by a causal chain of previous "water"-term usages back to the original dubbing of water. In the case of Earth, this effectively meant that what "water" originally referred to is H<sub>2</sub>O, and the semantic content of "water" would be fixed as H<sub>2</sub>O. The upshot of this is that Block's "water"-term on Twin Earth, where water was composed of XYZ, would never correctly refer to the XYZ or Twin-water, as the original Earth-dubbing of his "water"-term was fixed to H<sub>2</sub>O. This meant that to correctly refer to XYZ, Block would have to use another term like "Twin-water", "twater", or something like that.

However, Kripke's original notion of causal-historical reference, particularly regarding the reference of proper names, came under attack, most notably from Evans (1973) with his "Madagascar" example. According to Evans, "Madagascar"

originally referred to a part of the African mainland, and only later, possibly through some confusion, came to refer to the island off the African mainland (Evans, 1973, p.216). According to Kripke's causal-historical theory of reference, the name "Madagascar" should still only correctly refer to the part of the African mainland that was originally dubbed with that name at the beginning of the "Madagascar" causal chain. However, it intuitively seems that "Madagascar" now correctly refers to the African island and not its original referent. The main point of this example and others like it, is that in some cases the reference of proper names seemed to switch to a person or entity that is different to the person or entity that was originally dubbed by the name in question at the start of the causal chain.

A modified causal-historical theory of referring was put forward by Devitt (1981), which tried to explain Evans' example (and others like it) by the notion of "multiple grounding". The idea is that a name is grounded in its bearer on various occasions other than the original dubbing of the name. These other groundings might well be perceptual in nature and if there were enough of them, then the relevant name may well switch its reference from the original bearer to a different later one (Lycan, 2000, p.64). Devitt's notion is that a name is grounded in a bearer or referent by a network of causal chains, rather than just one linear causal chain stretching back to the original dubbing. For instance, with Evans' "Madagascar" example, it would be claimed that if there were enough perceptual groundings of the name "Madagascar" which refer to the African island, then the referent of the name Madagascar would change to the African island (Lycan, 2000, p.64).

With the above quotation, even though Block is referring to the natural-kind term “water” rather than a proper name, it seems that it is Devitt’s idea of multiple groundings that he has in mind (he also mentions Devitt (1981)) when speaking of “water” switching its reference on Twin Earth. Block speaks of his “water” being “grounded in the “dominant causal source”” of Earth, and presumably his “water” switches reference to twater after many years of grounding in the dominant causal source of Twin Earth. The first point to note here is that Block seems to be far more expansive in allowing his “water” to switch reference after many years on Twin Earth than does Fodor. Indeed, Fodor does not really address this issue at all, but speaks of the twins just being in the same context, say Twin Earth, and of then being in the same causal relationship to XYZ which means that the truth conditions of their water-thoughts will be the same. Moreover, the way that Fodor talks about the Earth twin’s “water” being “anchored” to H<sub>2</sub>O and the Twin Earth twin’s “water” being “anchored” to XYZ, one wonders whether Fodor ever intends a switching of water terms to occur when both twins are in the same context.

Nevertheless, the most important point here is that Block does allow that a natural-kind term such as “water” can switch its reference. Also of importance is that the timescale Block gives for the switching of reference to occur seems somewhat vague, and this could possibly be exploited. After all, Block claims that it is after “many years” on Twin Earth that his “water” would switch reference to twater. But how many years is this exactly, and what if Block had a large number of perceptual groundings of “water” on Twin Earth in a short period of time, could the reference switch then occur in matter of months, or weeks? Indeed, could the “water” reference

switch be claimed to occur after someone had thrown a bucket of XYZ over Block, as perhaps this would count as a direct or immediate perceptual grounding?

Bearing these points in mind, now imagine the following scenario. On Earth, I am seated at a table on which there is a jar containing some mud (earth and water). On Twin Earth, Twin-me is also seated at a table on which there is a jar containing some Twin-mud (earth and Twin-water). While seated at our respective tables, we are both having the thought that mud makes a good shampoo. There then occurs a bizarre and rather improbable event, I disappear from the Earth and instantly reappear on Twin Earth, at the very table my twin is sitting at. In his paper, Block describes a scenario where he just appears as he is in a counterfactual context, and claims that:

This specification of the mapping raises many familiar problems about the obscurity of counterfactuals...What are we supposed to think about the physics of a world in which someone (me, of all people) can appear on a street corner out of thin air? (Block, 1991, p.38).

However, I don't understand what Block is getting so worked up about here. He talks of himself appearing out of thin air, "me, of all people"; but if the actual situation is centred on him, surely it would be more scary if it was, say, Tony Blair, who appeared out of thin air in the counterfactual, rather than himself. As for the physics of the world concerned, this might be okay. I'm not entirely certain, but as quantum physics is usually seen as dealing with the probabilities of events occurring, there would be a probability attached to the event of someone disappearing from their world and reappearing on another world. Of course, the improbability of me disappearing from the Earth and reappearing on Twin Earth next to my twin is astronomically high, but it does not mean that an event of that kind is impossible. Therefore, I think that the

physics of a world would be okay with such an appearance out of thin air of either Block or myself.<sup>13</sup>

Therefore, the resultant situation is that I am on Twin Earth with my twin and we have a jar of Twin-mud in front of us. After the initial shock of my appearance has waned, Twin-me points at the jar and says “I was just thinking that mud makes a good shampoo”. I look at the jar of Twin-mud and say “What a coincidence, I also think that mud makes a good shampoo”. Now, recall that what we want here is there only to be a context of utterance/thought involved with my “mud”-thought, which means that we want my “mud” utterance/thought token to immediately refer to the Twin-mud, rather than still referring to Earth mud for a certain period of time. So, how can the intuitions involved here be turned so that this desired outcome sounds even remotely plausible?

Well, consider the following points. In his quotation, Block does acknowledge that the referent of his “water” would eventually switch to twater after many years. But what happens in those intervening years that is different to what happens in the first instant that Block sees twater? To me, the answer is that all that happens is Block builds up a history of causal and perceptual interactions or groundings with the dominant causal source, twater. But in the first instant that he sees twater, he also causally and perceptually interacts with the dominant causal source twater, couldn't this instantly ground twater as the referent of his word “water”. The only thing left out of this picture would be the time-period of vague specification, during which only more of the same interactions were experienced, which then eventuated in the long-awaited reference-switch of “water” to twater.



When I have the thought which I express with the utterance “mud makes a good shampoo”, whilst looking at the jar of Twin-mud in front of me, my “mud” thought/linguistic-token has the same causal relationship to the Twin-mud that Twin-me’s thought/linguistic-token has. This is of course what Fodor originally claimed, and presumably it is this causal relationship which entails that Twin-me’s “mud” is referring to Twin-mud and has the semantic content of Twin-mud or XYZ-mud, etc. However, the difference is that whereas Twin-me’s “mud” is correctly referring to Twin-mud, my “mud” is considered to be incorrectly referring to Twin-mud, due to Twin-me having a history of causal relationships with Twin-mud, a history which I do not possess.

But consider the intuitions about this claim. The first thing to note is that with the mapping theory, when me and my twin are both in the same context, it is the same conceptual and perceptual stereotype of the wide content, that our thoughts are to be mapped onto, as well as the same wide content. That is, it is the conception and appearance of the wide content that guides our thoughts to be mapped onto the wide content that they are. This means that with the mapping theory, it is acknowledged that the wide content of, say, mud on Twin Earth, may well be different to the wide content of mud on the Earth. This is indeed the case, when I am on Twin Earth, the wide content of the muddy-stuff that my “mud” is referring to is different, but the reference and mapping of my “mud”-thought is constrained in that context by the intensional-function of my “mud”-thought. That is, on Twin Earth, my “mud”-thought refers to, and locks onto, the muddy-stuff called “mud”, which has the same conceptual and experiential properties that it has on the Earth. Recall that it was these

conceptual and descriptive properties which were helping to individuate my “mud”-thought, not only on the Earth, but also on Twin Earth. Therefore, it seems to me that it is plausible to claim, despite the lack of a causal history, that my “mud”-thought/utterance does correctly refer to Twin-mud in the context of Twin Earth.

To test this intuition consider the following situation. On Twin Earth, mud, i.e., earth and H<sub>2</sub>O, does exist, but on this planet it has a blue crystalline appearance and is used in the making of jewellery and ornaments. Imagine that on Twin-me’s table, along with the jar of Twin-mud, there is also a candle-holder made of this blue crystalline Earth-mud. Now, is it plausible that my “mud”-thought will refer to, and get locked onto, the Twin-mud in the jar or the blue crystalline Earth-mud candle-holder? To me it seems totally counterintuitive to claim here that my “mud”-thought will refer to the Earth-mud candle-holder, as this is the substance that my “mud” has been in a causal-historical relationship with on the Earth. The more plausible position is that due to the conceptual and perceptual individuation conditions of my “mud”-thought, even though the wide content of Twin-mud is different to that of mud on the Earth, my “mud”-thought should correctly refer to, and map onto, the Twin-mud. The result of this is that the semantic content of my “mud”-thought/utterance will change immediately to be the same as the semantic content of Twin-me’s “mud”-thought/utterance, due to the constraining factors that were built-in to the intensional-function of my mud-thought.

What I am arguing for here can be summed up as follows. The individuation conditions of natural kind thoughts and utterances are normally and plausibly dependent on the “essence” that the particular natural kind has on the Earth (or in the

actual world); where this essence is then rigidified for that natural kind and so individuates the natural kind thoughts/utterances on other planets or across possible worlds (or at least does this for an indeterminate period of time according to Block). However, with the mapping theory, when thoughts and contexts are being mapped onto truth conditions or wide content, I have argued that it is not the essence of the natural kind which the thoughts are about which individuates them. What is rigidified from the Earth (or the actual world), is the conceptual and perceptual stereotype that the natural kind has there. When the natural kind thought is mapped onto wide content on Twin Earth, it seems plausible that what individuates that thought in that context, and which helps to lock the thought onto the appropriate wide content, are the rigidified conceptual and experiential properties of the intensional-function for that thought. The upshot of this is that the truth conditions that the intensional-function maps the natural kind thought onto, may well be different to the truth conditions that the thought previously had on the other planet or possible world. Indeed, this is the case with me and Twin-me above: when I suddenly appear in the Twin Earth context, the wide content of my “mud”-thought will be Twin-mud (dirt and XYZ), whereas, on the Earth the wide content of my “mud”-thought was mud (dirt and H<sub>2</sub>O), and vice versa for Twin-me appearing in the Earth context.

### 5.3.2 *Possible Problems Relating to My Construal of an Intensional-Function*

Now to Block’s second objection to my modified account of Fodor’s mapping theory of narrow content. This objection is really a number of interrelated criticisms of using

conceptual and perceptual properties of thoughts to constrain their mapping to wide content on Twin Earth.

### Wide Conceptions

For instance, in relation to the mapping of thoughts and utterances, Block claims that one could:

...require that the word “water” be used in each context with the same *wide conception* (in the sense of set of beliefs) that is associated with it in the actual world. This is, the mapping is the one that takes input contexts in which “water” is associated with the conception of being the local colorless odorless thirst-quenching liquid and maps them onto the obvious satisfaction conditions (Block, 1991, pp.41-42).

This sounds reasonable, but Block has some bad news for advocates of this strategy:

...once this proposal is made, it is immediately obvious that it too is doomed. The problem is this: as I have emphasized, what we intuitively want from a narrow content mapping is that it be one in which ‘water’ is used in each context with the same *narrow content* with which it is used in the actual world. Of course, we can’t require this explicitly, and so we must try via some other means...The means chosen in the proposal at hand is to require that ‘water’ be used in each context with the same beliefs about it (that is, beliefs with the same wide content) as in the actual world. There are a number of problems with this proposal, but the killer is that by requiring the same wide conception, we make what is supposed to be a narrow content *non-narrow*...there could be a physical and/or functional duplicate of me who does not share my wide ‘water’ conception. Of course, being a duplicate of me, he will have to share my ‘water’-*description* – colorless, odorless, thirst-quenching, liquid found in rivers and lakes. But he can live in a world in which, because of differences in his language community and his physical environment, none of these words in the description have the same meaning for him that they have for me. Indeed, he needn’t even *have* the concepts of colorlessness, odourlessness, liquid, etc...he might live in a Burgean language community which possesses concepts rather like our concepts of colorlessness, odourlessness, liquid, etc., but not those very concepts. Or, to use an example

from White (op.cit.), he may use 'liquid' to pick out a slippery granular solid that has the superficial appearance of a liquid (Block, 1991, p.42).

What can one say about the above? Firstly, it is fair to say that the strategy Block describes in the first quotation is pretty analogous to the constraining strategy that was earlier employed with the intensional-function of the mapped thought, in that it uses descriptive or perceptual factors to do the constraining. In the context of Earth or Twin Earth, either my or Twin-me's "water"-thought tokens would be mapped onto the wide content which was the local, colourless, odourless, thirst-quenching liquid in my environment. With the second larger quotation, Block seems to be claiming that even though my physical/functional twin would share with me a narrow "water"-description of the above-mentioned properties of water, my twin might not have the same wide "water" conception that I had. That is, this strategy of constraining the mapping seems to involve trying to keep the narrow and wide contents of me and my twin's thoughts the same, which may be possible with the narrow contents, but the wide contents could still differ. Therefore, the water-thought tokens would not be mapped onto the same wide contents, and so according to the mapping theory, they would not share identical narrow contents.

However, it seems fairly obvious that when Block claims that twins would share a water-description, but that due to differences in the language community of one twin, the meaning of the words that made up that description may mean something completely different to that twin, then we are back to Block's main objection against the mapping theory narrow content, viz., that it collapses into syntax. What Block is basically claiming is that my twin and I would only share the syntactic shapes of the words making up the water-description, i.e., odourless, colourless, liquid, etc. But of

course, it was to avoid this very result that we originally included conceptual and experiential properties in the individuation conditions of the particular thought I had on the Earth. For instance, if I have the thought which I express with the utterance “water is wet” on the Earth, then the intensional-function of this thought would not only consist of the syntactic properties of the thought, but would also include the conceptual and perceptual properties associated with that thought. That is, it would include my concept of water as being the local, odourless, colourless, thirst-quenching liquid, which fills the oceans and lakes, etc. Now, as far as I am concerned, not only is this a narrow water-description, but it is also a narrow water-conception. Of course, it can be argued that the narrow content of my water-thought does initially come from outside my body, from the external world; but I would argue that once the conceptual and perceptual input from the world is “internalised”, so to speak, then the narrow intensional-function of my water-thought can be fixed. The upshot of this would be that an internal physical/functional duplicate of myself, that is, a twin, would share the narrow water-description and water-conception that were fixed in the intensional-function of my thought that “water is wet” on the Earth.

Indeed, it makes me think that if there occurred a situation like the one Block envisages, where, say, Twin-me’s meaning for “thirst-quenching” is different to mine, then it could not be claimed that we are functional duplicates. As I understand Block, I think he is using “functional” to mean the same as “conceptual”, in that the functional or conceptual roles of me and Twin-me’s thoughts would be isomorphic, i.e., the way that me and Twin-me’s thoughts interacted with our sensory inputs and our other thoughts, which then produced behavioural outputs, would be identical. However, it seems clear that if Twin-me’s “thirst-quenching” meaning is distinct from

mine, then we are not going to be functional/conceptual duplicates. For consider, if I am thirsty and have the belief that there is a glass of water in front of me, then my knowing that water is thirst-quenching would, *ceteris paribus*, result in me drinking the water. However, if this same situation faced my thirsty twin, who did not have the same meaning or conception of “thirst-quenching” that I had, then *ceteris paribus*, it is not at all clear that he would drink the water in front him. Therefore, it seems to me that Block cannot claim that me and Twin-me are functional or conceptual duplicates, whilst also claiming that Twin-me’s “wide ‘water’ conception” is different to mine.<sup>14</sup>

#### A Conception of Narrow Content That Depends on Another Conception of Narrow Content

A further point that Block makes in connection to the functionalism issue is that the mapping theory of narrow content would not seem such an “attractive” proposition if it was seen to be dependent on another theory of “narrow content”, i.e., functionalism (Block, 1991, p.43). The first thing to say about this is that I have certainly used conceptual properties as part of the individuation conditions of the thoughts to be mapped, and conceptual role individuation of thoughts is indeed another theory of narrow content. The conceptual properties (along with perceptual properties) are used to constitute the mechanism of the “function” that is responsible for the same mapping of thoughts and contexts onto truth conditions. Moreover, as it is the “function” which Fodor takes to comprise the narrow content of the mapping theory (Fodor, 1987, p.53), then it does seem that Block is quite correct in his claim that mapping theory narrow content is dependent on another theory of narrow content.

Indeed, I think that one could also claim that as perceptual properties also play a role in individuating the mapped thoughts and part-comprising the mechanism of the “function” involved, then yet another theory of narrow content is involved with the mapping theory. That is, using the perceptual properties of the thought could be seen as similar to Fodor’s (1981) idea of the narrow “phenomenological belief”, the inspiration for which probably came from Putnam’s (1975) concept of the “stereotype” of a natural kind thought, e.g., about water, which Putnam considered would be the same for the twins involved in his Twin Earth thought experiment.

However, I don’t think that just because the mapping theory is dependent on other narrow content conceptions, that this fact necessarily means that the mapping theory is less attractive. Recall that one of the aims of Fodor’s mapping theory was to ensure that the twins’ thoughts would have the same truth conditions or wide content in the same context. This in turn led to Fodor’s extensional identity criterion for mental contents, i.e., “Two thought contents are identical only if they effect the same mapping of thoughts and contexts onto truth conditions” (Fodor, 1987, p.48). This then led to the extensional identity criterion for narrow content, which is that the narrow content of two thoughts is that function (or thought-content) of the two thoughts which effects the same mapping of thoughts and contexts onto truth conditions. The upshot of all this is that it is surely not surprising that the narrow function shared between two thoughts which effects their same mapping onto truth conditions, is constituted by conceptual and perceptual properties; even though these properties are also utilised as conceptions of narrow content in their own right. For the twins’ thoughts to be mapped onto particular wide contents in the external world,



this surely involves, and depends on, the twins' relationship to objects and events in the external world; that is, how they conceive and perceive that external world.<sup>15</sup>

Therefore, in this chapter Fodor's narrow mapping theory has faced very specific and detailed criticism, particularly from Block. However, I have tried my best to deflect this criticism by expanding on certain aspects of Fodor's conception, e.g., by trying to figure out exactly what a "function" is and the role it plays in the mapping theory.

#### 5.4 PART 2 CONCLUSION

This part of the thesis has been concerned with a construal of narrow content which is based on Kaplan's notion of "character", which comes from his influential work on the semantics of indexical expressions. The main difference with Kaplan's notion is of course that Fodor's narrow content is concerned with thoughts or propositional attitudes, rather than linguistic expressions. This certainly does complicate matters in relation to deciding what actually determines the narrow content of those thoughts.

With the character of an indexical expression such as "I", the character is straightforwardly determined by linguistic convention, so that it refers to the person who uttered, wrote, etc., the particular "I" in question. From this, it is again straightforward to see how the character can act as a function from contexts of utterance to the content that the indexical expression would have, i.e., if I had uttered the "I", then the content of the "I" would be MD.

However, when Fodor claims that the narrow content of a thought is a function from contexts onto wide contents, it is not as straightforward to see how the function in question is determined. This fact is exploited by Block's objection that Fodor's mapping theory narrow content collapses into syntax. The crux of Block's claim is that my twin's context of acquisition for the thought/utterance in question could be different to my context of acquisition for the thought/utterance in question. The result of this is that my twin's particular thought/utterance could be mapped onto a completely different wide content in his context, to that of which my particular thought/utterance is mapped in my context. I claimed that this move by Block is unfair to Fodor, as in standard Twin Earth thought experiments, the contexts of acquisition for the thoughts concerned are kept the same, in order to gauge what effect the subjectively imperceptible differences in the physical and/or social environments of the twins having the thoughts would have on the contents of those thoughts.

With "contexts of acquisition" here, these would include the conceptual and experiential elements involved in acquiring the thoughts. Personally, I think it is the same situation with Fodor's mapping theory of narrow content, though of course here we are trying to gauge what function (or "content") the relevant thoughts share, despite the subjectively imperceptible differences in the physical and/or social environments of the thoughts. Of course, Block would probably claim that what we are really doing here is keeping the narrow content of the thoughts the same. But I would deny this, all that we are doing is applying the same thought experiment criteria to Fodor's mapping theory thought experiments that are applied to the thoughts experiments of Putnam, Burge, etc.

Allowing the contexts of acquisition for the relevant thoughts to be the same would also help to avoid the rather nasty problem that the causal-historical theory of reference caused for Fodor's mapping theory. Indeed, it seems that the causal history of the concepts that make up thoughts will be problematic for any theory of narrow content based on Kaplan's character. At least, if that narrow content theory just imagines a person having a context of thought for the relevant thought. In this situation, there will always be intuitions that the person is not correctly referring to the object/person now in front of them in their present context, but are rather still referring to the object/person that they have had a significant causal history of interactions with, in the previous context. Of course, the mapping theory could still use just contexts of thought, if my attempts to deflect the causal-historical intuitions were considered successful. However, it seems more plausible to me for the contexts of acquisition to be the same. If this is allowed, then the causal-historical intuitions should be nullified, as when I have been in my twin's context of acquisition "for some period of time", then the intuition would be that my "water" thought symbol is correctly referring to the same wide content as my twin's "water" thought symbol, i.e., XYZ.

However, it must be noted that Fodor's mapping theory narrow content is too coarse-grained to register some narrow distinctions of content, i.e., the claim that two thoughts share the same narrow content if they are coextensional in a context does seem to generalise over finer content-distinctions that could be made. For example, the subjects concerned may have thoughts made up of distinct conceptions of the same wide content, which may differently affect their respective behaviour.

Nevertheless, I think it should be borne in mind that Fodor had a number of aims in

his 1987. He wanted a narrow content that could be used by a scientific psychology, but he also wanted to repair the connection between thought-content identity and extensional identity, a connection that he believed had been broken by Putnam's Twin Earth thought experiment. Fodor believed that he had repaired this connection, at least in a context, with his extensional identity criterion for thought-contents, and also for narrow contents.

Nevertheless, we now come to the important question of how Fodor's mapping theory fares in satisfying the three conditions of adequacy for narrow content. This is going to be difficult to figure out, as I found Fodor's position somewhat confusing. It is plausible to take the view that the mapping theory narrow content does not satisfy condition (1), i.e., it does not provide a narrow content that is considered to be internal, "in the head", and is shared between the duplicates. The reason I say this is that Fodor takes the view that his narrow content is only "potential" content, it only becomes proper belief content when it is anchored to a context. Moreover, when it is anchored to a context, it is not narrow content anymore, but is now wide content. The mapping theory narrow content does satisfy condition (2), in that even though Fodor claims the narrow content cannot be expressed, it can be given an abstracted specification. With condition (3), it is again plausible to take the view that the mapping theory does not satisfy this condition, as it is not the narrow content which provides the psychological generalisation that subsumes the behaviour of the duplicates in the same context. The behaviour of the duplicates is subsumed by a psychological explanation that appeals to the shared wide content in the same context, i.e., H<sub>2</sub>O in the Earth context and XYZ in the Twin Earth context. Moreover, the mapping theory is also not flexible enough to deal with Frege cases. Therefore, even

though Fodor's narrow mapping theory is a brave effort, it does not satisfy all the conditions of adequacy for narrow content.

## Notes to Chapter 5

<sup>1</sup> It is Block who first refers to Fodor's conception of narrow content as the "mapping theory" in his 1991 article.

<sup>2</sup> This distinction in a "context of acquisition" was pointed out to me by my supervisor.

<sup>3</sup> Block's charge of narrow content collapsing into syntax seems fairly similar to an objection raised by Segal (2000) against using the notion of character to specify the narrow content of thoughts. When considering the addition of meta-linguistic descriptions into the notion of character, Segal thinks that this move would be problematic, as most "...concepts are deferential...Narrow content then ends up being almost exclusively about words. Consider "arthritis". Is its character merely "thing called 'arthritis' in my home environment"? If so, then this clearly does not suffice to determine cognitive content. Consider a cousin Earth in which "arthritis" means *water*. Then, on this proposal, the cognitive content of a normal Earth individual's concept of arthritis is the same as that of a normal cousin Earth individual's concept of water" (2000, pp.114-115). So here, it seems that the idea that the concept that a word expresses is deferential to the word's "home environment", is equivalent to the idea that a word's "context of acquisition" determines which concept it is expressing in that context. Therefore, if a particular word expresses a different concept in another home environment or context of acquisition, then in both cases, the only things that the words will have in common are their syntactic shapes. My following reply to Block's objection can also be taken as a reply to Segal's objection.

<sup>4</sup> It is the lack of conceptual properties in Fodor's notion of the narrow function of a thought that Loar (1988a) criticises on p.197.

<sup>5</sup> Stalnaker (1991) seems to make this suggestion on p.233.

<sup>6</sup> Perhaps one could say that the intensional function of a particular thought determines which properties an object in the world must possess in order to be included in the extension of that thought. For discussion on this topic see Searle (1983), Horowitz (1996) and Jackson (1998).

<sup>7</sup> I don't think this presupposition is problematic, as Block himself seems quite happy to adopt the LOT hypothesis in his paper (1991, pp.34-35).

<sup>8</sup> This notion is taken from Lycan (2000) p.155 where he is discussing intensional semantics.

<sup>9</sup> Block (1991) uses the idea of satisfaction conditions for the thoughts on pp.41-42.

<sup>10</sup> I utilised both Block (1991) and Stalnaker (1989) for this mapping formulation of Kaplan's idea of character.

<sup>11</sup> In saying this, I don't mean to claim that the mapping theory could *only* work with a context of thought. It could work, with a lot less (causal-historical) trouble, with contexts of acquisition and thought. But of course, it would have to be allowed that we could stipulate that the Twin Earth context of acquisition would be identical to the Earth context of acquisition (except for the XYZ of course).

<sup>12</sup> With the following potted history of the causal-historical theory of reference and objections to it, I have been greatly helped by Lycan (2000) pp.60-68.

<sup>13</sup> Moreover, there is a philosophical tradition of using thought experiments such as the above, involving strange and improbable events, to make certain points, raise philosophical problems, etc. (Thanks to my supervisor for this point).

<sup>14</sup> I think this also addresses an objection to Fodor's mapping theory that Manfredi (1993) makes, where he claims that, contra Fodor, "Two individuals in a given context, with very similar learning histories, could have thoughts with identical broad content without sharing identical causal powers" (1993, p.7). The crux of this claim is that Manfredi imagines a situation where on the Earth, whenever he sees water, he thinks that it is good to drink, and when he does this, he is in a brain state of type P. However, Manfredi claims his twin on twin-Earth could be in a brain state of type P and yet not believe that twater is good to drink (1993, p.7). However, it seems to me that in a situation like this Manfredi and his "twin" are not physical/functional duplicates as this is understood in the standard Twin Earth thought experiments. If Manfredi and his twin are physical/functional duplicates, then when the former is in a brain state of type P and thinks that water is good to drink, then, by hypothesis, his twin is also going to be in a brain state of type P and will also think that twater is good to drink.

<sup>15</sup> As Fodor himself declares, "...what *has* narrow content is not a syntactical object as such, but a syntactical object *together with the bundle of mechanisms that mediates its mind/world relations*" (1991b, p.269). By "syntactical object" I take Fodor to be referring to the particular Mentalese thought-token involved.

# **PART 3**

## **NARROW REPRESENTATIONAL CONTENT**

In this part of the thesis a representational construal of narrow content will be examined, the exemplar of which is Dennett's conception of notional worlds. In chapter 6 the notional worlds conception will be explicated and assessed. Criticisms of it will also be made, particularly from Loar, whose critique leads me to have an idea of a more preferable way to formulate a subject's notional world. In chapter 7 I give what I call a narrow representational reformulation of Dennett's notional worlds, which uses the idea I had in the previous chapter. In doing this I utilise an unwieldy thought experiment involving ufos and aliens, with the aim of specifying what narrow representational content is shared between duplicates in a Twin Earth-type scenario.

## CHAPTER 6

### DENNETT'S NOTIONAL WORLDS

Dennett's (1982) idea of the "notional world" of an agent is his attempt to answer the question, "what is the organismic contribution to the fixation of propositional attitudes" (Dennett, 1982, p.36); or, put another way, what contribution do the internal physical states of the agent make to the content of that agent's beliefs and other intentional states. The organismic contribution could then be considered to constitute the narrow content of the agent's beliefs. Dennett's objectives for his notional attitude psychology are the following:

We want it to work out that I and my *Doppelganger* – and any other narrow-psychological twins – have exactly the same notional attitudes, so that differences in propositional attitudes are due entirely to the different environmental contributions. But we also want it to work out that you and I, no psychological twins but "of like mind" on several topics, share a variety of notional attitudes (Dennett, 1982, p.38).

To explain what Dennett means by the above quotation and with notional worlds in general, consider the following example. Imagine that I jump into a swimming pool on the Earth, and have the belief that the water is very cold. Now, imagine that my doppelganger on Twin Earth also jumps into a swimming pool and also has the belief that the water is very cold. Both situations are considered identical except for the usual difference, which is that on Twin Earth water is composed of XYZ molecules rather than H<sub>2</sub>O molecules. If the externalist conclusions which are drawn from



Putnam's Twin Earth thought experiment are accepted, this one difference is taken to show that my doppelganger and I's respective beliefs involve different propositions and so are different propositional attitudes. My belief expresses the proposition that the water is very cold, while my doppelganger's belief expresses the proposition that the XYZ or Twin-water is very cold. From the above quotation, it seems that Dennett does accept the externalist conclusions, and holds the view that differences between me and my doppelganger's propositional attitudes are due to the different wide content of those beliefs.

However, despite these wide differences, it could be claimed that my doppelganger and I both share the same notional world and notional attitudes within that world. In our respective notional worlds, it seems plausible to claim that we are sharing the same notional experiences, which contain the same notional events and objects. For example, we both have qualitatively identical perceptual and phenomenal experiences in our respective notional worlds, of having jumped into our respective swimming pools, and, I suppose Dennett would claim, that we are also having the same respective notional water-attitudes. By this, I think Dennett means that me and my doppelganger would both have the same narrow notional attitude that the clear and transparent liquid which fills the swimming pool, and is called "water" by our respective social communities, is very cold. The narrow content of me and my doppelganger's shared notional attitudes is something which could not be used to differentiate between the worlds we could be located in, i.e., from our notional-water attitudes it could not be decided that we were on Earth rather than Twin Earth or vice versa.

The idea seems to be that an agent's notional attitudes are the step before his or her wide propositional attitudes are fixed. The latter are of course fixed by actually denoting the objects and events in the external world that are then contained in the agent's propositional attitudes. Therefore, it could be claimed that notional attitudes are like propositional attitudes, except that the former do not have a form of reference that actually fixes or anchors them to the events and objects that are in the real world. Perhaps it could be said that notional attitudes are world-directed or intentional, in that they are about events and objects in the real world, but that the causal chains which connect them to the external world are only connected so far, that is, to the perceptual appearances of events and objects to the subject concerned.<sup>1</sup> Dennett claims that some notional objects may have counterpart objects in the real world (Dennett, 1982, p.38), the upshot of this is that some notional attitudes about notional objects could have counterpart propositional attitudes in the real world. With me and my doppelganger, our respective same notional attitudes do have counterpart propositional attitudes, but, of course, the latter are not the same.

As I have described the ideas of Dennett's notional worlds and notional attitudes up to this point, it seems quite natural to come to the conclusion that the formulation of an organism's notional world is going to take place from the subjective or first-person perspective of the organism concerned, i.e., the notional world will be formulated from the organism's subjective sensory experience, how the world is represented by the organism itself.<sup>2</sup> Indeed, Dennett himself admits that his notional objects might be considered to be the "intentional objects" of Brentano and that his notional worlds are similar to Husserl's idea of "bracketing" (Dennett, 1982, p39).<sup>3</sup> However, this where Dennett makes an interesting move with his idea of notional worlds. Rather

than employing the first-person “auto-phenomenology” of Brentano and Husserl in formulating an organism’s notional world, Dennett considers that an organism’s notional world can be formulated from the objective third-person perspective of another. This process of “hetero-phenomenology” (as Dennett calls it) means that “[a] notional world should be viewed as a sort of *fictional* world devised by a theorist, a third-party observer, in order to characterize the narrow-psychological states of a subject” (Dennett, 1982, p.38). This move of Dennett’s seems to stem from his suspicions of the subjective realm, for instance, at one point he claims that the idea of a notional world which is “...constituted by the mind or experience of a subject...may be nothing more than an eternally tempting mistake” (Dennett, 1982, pp.51-52). Therefore, how does Dennett aim to formulate a subject’s notional world from an objective third-person point of view?

### **6.1 How Dennett Proposes to Formulate a Subject’s Notional World**

To begin with, Dennett speaks of using evolutionary/biological factors for formulating the notional world of an organism, an organism which he imagines is still alive, but frozen, and so is “cut off from any environment” (Dennett, 1982, p.41). Ex hypothesi, by examining the organism’s internal structure, behavioural dispositions and biological needs, we can formulate its notional world or “...the environment (or class of environments) for which the organism as currently constituted is best fitted” (Dennett, 1982, p.42). Moreover, Dennett thinks that this method of notional world formulation can be carried out with organisms like humans, who have:

...internal structure and dispositional traits so rich in information about the environment in which they grew up that we could in principle say: this organism is best fitted to an environment in which there is a city called Boston, in which the organism spent its youth, in the company of organisms named...and so forth (Dennett, 1982, p.43).

Therefore, it can be seen that with this method of notional world formulation, by hypothesis, all the information needed is able to be read off from the physical properties and behavioural dispositions of the creature, from an objective third-person perspective. The notional world is then formulated to match this information, and so will be a world that the creature is best fitted to live in.<sup>4</sup> Of course, it should be noted here that with Dennett's method of notional world formulation, the subjective first-person perspective of, e.g., the human's, sensory experience, is not directly utilised. This ties in with Dennett's previously mentioned preference for hetero-phenomenology instead of auto-phenomenology, the latter of which would involve formulating the notional world solely from the subjective first-person perspective of the human's sensory experience, which Dennett viewed with some suspicion.

It is from the first-person perspective of the human that his or her beliefs would be expected to have a role in formulating that person's notional world. For example, using possible worlds semantics, we could say that the person's notional world was the set of possible worlds where the maximal number of that person's beliefs were true. However, Dennett claims that a person's notional world could "...be viewed formally as the set of possible worlds consistent with the maximal description" (Dennett, 1982, p.44). Of course, by the "maximal description", Dennett means the description which the third-person theorist gains from analysis of the physical constitution and behavioural dispositions of the person. That is, there are no beliefs, *qua* internal belief-states, involved in the formulation of the person's notional world.

Perhaps Dennett's choice of not using the person's subjective internal beliefs to help construct that person's notional world, comes from his instrumentalist conception of belief attribution. It could be viewed as instrumentalist in the sense that Dennett does not think that beliefs (and desires) actually exist, but that it is useful to attribute them to a system if it helps explain and predict that system's behaviour. Dennett calls this conception the "intentional stance". This is in contrast to the "physical stance", which predicts a system's behaviour by mainly concentrating on its physical constitution and applying physical laws; and the "design stance", which predicts a system's behaviour by viewing it as having a certain design, which will normally behave as it is supposed to, according to the design (Dennett, 1981, pp.340-341). Even though Dennett is of the opinion that beliefs (and desires) do not have an objective existence, he does believe that when adopting the intentional stance to humans, real objective behavioural patterns are observable and describable, which are not available when adopting the physical and design stances (Dennett, 1981, pp.344-345).

Be this as it may, Dennett acknowledges that his method of notional world formulation does depend on the assumption that the organismic contribution does ultimately depend on the physical constitution of the organism, or as he also calls it, "...the supervenience of (narrow) psychological traits on physical traits" (Dennett, 1982), p.44). Dennett also points out that from the above notional world, one would not be able to distinguish Boston from Twin Boston; however, this is to be expected, as we noted earlier, it seems a feature of narrow content that it does not discriminate between Earth and Twin Earth, or the actual and counterfactual Earth.

Therefore, it is by the above method of notional world formulation that Dennett hopes to find out what the organismic (i.e. narrow content) contribution is to the fixing of a subject's propositional attitudes.

## 6.2 Assessments and Criticisms of Dennett's Notional Worlds

### 6.2.1 *Loar's Views on Notional Worlds*

Having described some of the workings of Dennett's idea of notional worlds, it is now time to examine some of the views and criticisms of this idea. Loar's (1988a, 1988b) discussion of Dennett's notional worlds is interesting as he notes that a comparison has been drawn with the aforementioned notional worlds and Loar's own idea of "context-indeterminate realization conditions (CIRC)" (Loar, 1988a, p.197). It will be recalled that Loar's idea of CIRC was an attempt to circumvent the problem of narrow content not being truth-conditional, and so Loar believed, of lacking intentionality. Given an agent's narrow beliefs, obtaining the CIRC of the beliefs involves abstracting away from context, and considering a world in which the agent's beliefs are true and not misconceptions. One could then say that the CIRC of the agent's beliefs determined a set of possible worlds in which the agent's beliefs are true and not misconceptions (Loar, 1988a, p.108). It was by this indirect method that Loar attempted to endow narrow content with some truth-evaluability.

It can be seen immediately that there does seem to be a similarity at least between the set of possible worlds determined by the CIRC, and the set of worlds that are

determined in forming a person's notional world. With the CIRC we have a set of possible worlds where a person's beliefs are realized and true, and with notional worlds we have a set of possible worlds where a person's beliefs about objects and states of affairs are also realized, i.e., the notional world of a person is identical to the way the person believes the actual world is. Indeed, one could phrase it as saying that in a person's notional world his beliefs are maximally true.

However, Loar points out a number of important differences between his CIRC and Dennett's notional worlds. The most obvious difference is that while Loar's CIRC are a subjective first-person conception, Dennett's notional worlds seem to be an objective third-person conception (Loar, 1988b, p.n202). This leads on to the next difference which is that CIRC are formulated from internal belief-states, whilst, as we have seen, notional worlds are formulated from the maximal description of the person's internal structure and behavioural dispositions. Loar then claims the following:

His [Dennett's] third-person notional worlds are invoked (it seems to me) instrumentally; there are no inner (functional or neural, say) non-instrumentally ascribable facts about a person x which *constitute* the truth-conditions of an assertion that x's notional world is such and such (Loar, 1988b, p.n202).

Loar's claims seem to tie in with the conclusions that were reached in the previous section. That is, an agent's notional world seems to be mainly formulated from behavioural dispositions; with the notional beliefs and other notional attitudes being instrumentally ascribed directly from the dispositions. Moreover, Loar believes that because notional worlds cannot be objectively "determined by inner facts" (Loar,

1988b, p.202), this leaves Dennett's conception of narrow content open to attack

from the following externalist argument:

1. The psychological contents of thoughts (i.e., the individuation conditions of thoughts in commonsense psychological explanation) consist in representational properties;
2. Representational properties are not determined by internal factors;
3. Therefore, psychological contents are not determined by internal factors (Loar, 1988b, p.135).

I think Loar's reasoning behind the above argument is roughly the following.

Because commonsense psychological explanation involves, as Loar puts it, states that are said to have a certain content, or "...*how thoughts represent things*" (Loar, 1988b, p.134), then this would seem to involve states with referential truth-conditions, which, it is generally acknowledged, cannot be determined by conceptual role individuation or any other internal factors (Loar, 1988b, p.134). The conclusion is that "...internalism about psychological contents – the thesis that commonsense recognizes "narrow contents" – is incoherent" (Loar, 1988b, pp.134-135). It is of course by conceptual role individuation that Loar defines his narrow, internal, psychological content, so it can be seen that the above argument is a real threat to Loar's conception of narrow content. The way that Loar gets round the argument is by claiming that psychological explanation appeals to states that have "content-like" properties that only have the appearance of intentionality or representationality (Loar, 1988b, p.136). What Loar is admitting to here is that his psychological content is not really content as it is normally understood, i.e., a content which involves referential truth-conditions. Moreover, he also claims that his psychological content only has this appearance of representationality from the subjective perspective, i.e., if one moved to an objective perspective, the appearance of representationality would not be



observable (Loar, 1988b, p.136). The upshot of all this is that with the above argument Loar can claim that either premise (1) is false if it only refers to properties that are representational from an objective third-person perspective (Loar, 1988b, pp.136-137); or that premise (2) is false if it does not take into account the internally determined content-like properties of the psychological content of thoughts.<sup>5</sup>

In describing the method that Loar uses to counter the externalist argument, I have left out one important point; namely, how Loar proposes to give an objective, non-intentional explanation of the content-like and apparently representational properties of a thought's psychological content. Or to put it as Loar did above, how his narrow content can be objectively determined by inner facts. This is necessary for Loar, otherwise his claim that premise (2) is false could be doubted, i.e., it could be doubted that the factors which are responsible for the quasi-representational properties of his psychological content are indeed *internal* factors. If I understand Loar correctly, his basic answer to how his narrow content can be objectively determined by inner facts, is that an objective view can be taken of the conceptual role individuation of thoughts that is available from the subjective perspective (or as Loar calls it, the introspective/projective perspective) (Loar, 1988b, pp.136-137). The upshot of this is that Loar believes his narrow psychological content can avoid the above externalist argument, due to the fact that his narrow content can be objectively determined by internal properties (or inner facts). However, it is this strategy which Loar thinks is not available to Dennett's narrow notional worlds in dealing with the externalist argument, due to the instrumentalist nature of the conception.

### 6.2.2 *A Clarification of Dennett's Notional Worlds Conception*

But is Loar correct in his analysis of Dennett's notional worlds? Indeed, is there the possibility that Loar (and myself) have not done full justice to Dennett's conception of notional worlds? To ascertain whether this is the case, let us re-examine some of the issues involved with notional worlds. From my reading of Dennett, I think that it is accurate to say that the method of formulating an agent's notional world is an objective third-person process. It is only the internal structure and dispositional traits of the agent that are utilised in formulating the notional world. The internal belief-states of the agent, as apprehended from the subjective first-person perspective of the agent, are not utilised in the formulation. This is fair enough, as Dennett is after all trying to find out the organismic (i.e., physical) contribution to the fixing of an agent's propositional attitudes. If he thinks that belief-states don't have an objective (physical?) existence, then he is quite right not to include them in the organismic contribution, which in turn means that he can't utilise belief-states qua internal/subjective belief-states, in the formulation of an agent's notional world. This is the difference between Dennett's hetero-phenomenology and the auto-phenomenology of Brentano et al.

Therefore, having said all this, I suppose that Dennett could still plausibly claim that even though an objective third-person method of formulating an agent's notional world is used, the resulting notional beliefs that make up that world, could still be apprehended from the first-person perspective of the agent. That is, it could be claimed that a notional world did include subjective information, but that this information would be constrained by the objective information involved in the

formulation of that notional world.<sup>6</sup> So even though the notional attitudes of an agent may be instrumentally ascribed with the formulation of that agent's notional world, they would have an "existence" from the agent's first-person perspective. I think this is a fair point to make, as my previous description of Dennett's notional worlds may have given the impression that there was nothing happening from the subjective first-person perspective of the agent. It is perhaps more accurate to say that there is subjective information in an agent's notional world, but that this does not play a part in the actual formulation of the notional world.

Another point worth considering is how the word "objective" has been used by Loar (and myself) in the previous few paragraphs. There is an "ontological objectivity" which is concerned with the properties of things in the world, and there is an "epistemic objectivity" which is concerned with the properties of judgements. Perhaps Dennett could agree that belief-states did not have an objective existence (in the ontological sense), but he could still claim that one could make objective judgements (in the epistemic sense) about them.<sup>7</sup> I think this is correct, but would this epistemic objectivity be enough to enable Dennett to escape from the externalist argument that Loar mentions above? It seems to me that it might. I suppose we could try something similar to Loar's above strategy, where he claims that an objective (in the epistemic sense, I think) view of the pattern of conceptual role individuation of thoughts would provide the inner facts that determine his narrow psychological content. With Dennett's notional worlds, perhaps we could also take an epistemically objective view of the pattern of inferential relations between an agent's notional thoughts, which then would provide the inner facts that determined the narrow notional world of the agent. Indeed, this utilisation of the objective patterns of

conceptual role individuation of an agent's notional thoughts is perhaps what Dennett had in mind; as mentioned earlier, he does claim that his notional world psychology could be viewed as a version of a "narrow conceptual role semantics".

Interestingly, the idea of the objective conceptual patterns of the beliefs and thoughts in a person's notional world, seems slightly similar to Dennett's previously mentioned objective "real patterns" that are only observable when adopting the intentional stance towards certain systems, e.g., humans. However, it should be pointed out that when Dennett talks of real objective patterns, like a true instrumentalist, he is referring to patterns of behaviour among intentional systems like humans (Dennett, 1981, pp.344-345), and not patterns of the conceptual interactions among a person's beliefs and thoughts. Nevertheless, when one tries "projecting" one's thoughts or a notional world onto another person, in order to specify how they are conceiving things, it could be claimed that how that person is conceiving things will mainly be exemplified by their behaviour and dispositions to behave. The difference seems to be that Dennett just stays with the patterns of behaviour, whilst Loar would ascribe inner beliefs and thoughts which had a particular conceptual pattern or structure, and which were causally efficacious in explaining the person's behaviour.

Having said all this, does Dennett still have an ontologically objective option that would enable him to escape the above externalist argument? That is, are there actually existing internal properties or inner facts of an agent that could be said to determine his narrow notional world? As we have seen, Loar doesn't think that Dennett has this option available, as the way the notional world of an agent is formulated is from the objective third-person perspective mainly using his

behavioural dispositions. However, it is worth noting that as well as the agent's dispositional traits, Dennett also claims that the "internal structure" of the agent is also utilised in constructing the agent's notional world. Now, this internal structure is presumably physical in nature, indeed, it seems feasible to assume that it could be internal *neuropsychical* structure that Dennett is alluding to. If this is accepted, then it seems that Dennett does have the option of claiming that internal properties or inner facts do determine an agent's notional world. He can claim that an agent's internal neurophysiological properties contribute towards the determination of his narrow notional world. Therefore, it seems that Dennett's conception of notional worlds is again able to escape the externalist argument that Loar mentions.

Nevertheless, I can't help feeling that there is something odd or unintuitive about the objective third-person way that Dennett constructs his notional worlds. As I said above, I can understand why he has taken this route, in that he wants to know only what the organismic contribution is to the fixing of an agent's propositional attitudes. Moreover, he also seems intent on avoiding the subjective first-person realm of the agent's experience, in formulating the notional world. However, is there a more plausible method to formulate an agent's notional world, which perhaps uses the internal belief-states of the agent, while still keeping to the spirit of Dennett's objective third-person approach? I think there might be a possibility of doing this, but this idea will be examined in detail in the next chapter.

### 6.2.3 *Notional Worlds and Representational States*

In this section I will briefly consider some criticism of Dennett's notional worlds by Stalnaker (1989), as this has a bearing on what I will discuss in the next chapter.

Stalnaker wonders what it is that differentiates the notional worlds of narrow notional attitudes from the possible worlds of ordinary, wide propositional attitudes (Stalnaker, 1989, p.181). If me and my doppelganger are supposed to have the same notional world, even though we live in different real worlds, then, as Stalnaker claims, "The set of notional worlds that define the narrow contents of a person's beliefs is something like the worlds that *are* the way that the person takes the real world to be" (Stalnaker, 1989, p.181). If described like this, notional worlds do seem very similar to possible worlds, as with the latter, there is the actual world, and then non-actual or fictional possible worlds, which characterise the wide contents of beliefs (Stalnaker, 1989, p.181). There is also another similarity which Stalnaker observes between notional worlds and possible worlds, namely, that the narrow and wide content that is characterised in those respective worlds, will be propositional content (Stalnaker, 1989, p.181). So the original question can now be asked in a different way, i.e., what is it that differentiates narrow propositional contents from wide propositional contents? Stalnaker claims that the answer to this question:

...is to be found in the different answers that the two theories give to the question, "in virtue of what facts do a believer's beliefs have the (notional or propositional) contents that they have?" (Stalnaker, 1989, p.181).

I think that Stalnaker is correct with the above answer. Indeed, when one thinks about it, it has to be the correct answer, considering Dennett's original purpose for his notional worlds. He wanted to find out what the organismic contribution is to the

fixing of belief contents, which will be a function involving only the internal states and properties of the believer. This is in contrast to wide propositional attitude psychology, which sees the fixing of belief contents “as a function of relations between the believer and his actual environment (Stalnaker, 1989, p.181). The set of notional worlds that are picked out to characterise the narrow content of a belief will be ones which are only defined by the internal states and properties of the believer; or as Dennett puts it himself, a notional world could be described as “...the set of possible worlds consistent with the maximal description” (Dennett, 1982, p.44), the latter of which is produced by the third-person theorist, *ex hypothesi*, just from the internal structure and dispositional traits of the believer.

However, Stalnaker is not convinced that Dennett’s strategy will work. He claims that:

...the attempt to recover information about a virtual environment without making any assumptions at all about the actual environment is just too unconstrained to work. Imagine a purely internal description of the movements that I am disposed to make under various internal conditions, as I walk down the streets of Boston going places to satisfy my wants and needs, a description that makes no reference to what is going on either specifically or in general beyond my skin. How could anything about *Boston*, or about Boston-like cities, be recovered from such a description? (Stalnaker, 1989, p.185).

This seems a fair criticism from Stalnaker, especially if one considers trying to formulate Stalnaker’s notional world of Boston just from a description of his internal structure and dispositional traits. One would think that this was almost impossible without including in the description some information about things outside of Stalnaker’s skin. However, is Stalnaker correct to presuppose that an agent’s notional world could not be formulated with internal states and properties that were in some

sense world-directed? Stalnaker's use of the phrase "reference" above is ambiguous, as of course if the agent's internal states were *referring* to things outside the agent's skin, then this would mean that the intentional content of those states would be truth-conditional, and so the content involved would be wide. However, couldn't it be claimed that if the internal states of an agent consisted of some sort of quasi-representational content, as well as dispositional traits, then one's formulation of the agent's notional world would involve a conception of things outside of the agent's skin. This conception would not really be representational, it would not refer to objects in the actual world, but the conception would still be world-directed or intentional in the sense that it was "about" things that were outside the skin of the agent.

It seems to me that Dennett is a little vague as to whether representational states are included in his formulation of notional worlds. He mostly talks of internal structure and behavioural dispositions being involved in the formulation. Moreover, in Dennett (1987) he makes it sound as if notional world formulation is just a special case of radical interpretation (Dennett, 1987, p.209). This latter view involves a suitably placed and knowledgeable expert assigning intentional states to an agent on the basis of his or her behaviour in various situations (Braddon-Mitchell/Jackson, 1996, p.273). The behaviour in this case would also involve the utterances that the agent made, and Dennett supposes that expert Martians could induce the agent "...hypothetically, to recount in its native tongue as much of its biography as it could muster" (Dennett, 1987, p.209). All of this has a familiar instrumentalist ring to it, as Dennett seems intent on staying away from the agent's subjective experience and the possibility of inner representational states.



However, at other times, Dennett speaks of notional attitudes which have notional objects, and speaks explicitly of "...the subject's representations" (Dennett, 1982, p.41). Indeed, at one point Dennett claims that the "...idea of a notional world, then, is the idea of a model – but not necessarily the actual, real, true model – of one's internal representations" (Dennett, 1982, p.40). But there is even the possibility that Dennett's talk of an agent's "internal representations" might itself be ambiguous. For instance, Dennett (1981) claims that when successfully adopting the intentional stance towards a particular object, and one feels the urge to claim that some of the object's internal states may be internal representations, then "*[w]hat makes some internal feature of a thing a representation could only be its role in regulating the behaviour of an intentional system*" (Dennett, 1981, p.347). When Dennett makes this claim, he is actually talking about a thermostat, but one can see that this line of thought could end in some sort of dispositional result, when applied to the representational states of a human.

Therefore, in this chapter Dennett's idea of notional worlds has been explicated and assessed. I agreed with Loar that with Dennett's original idea, the formulation of a subject's notional world seemed to involve ascribing beliefs to the subject in an instrumentalist manner, which could be problematic and open to attack from certain externalist arguments. This gave me an idea for an alternative method of formulating a subject's notional world which will be described in detail in the next chapter.

## Notes to Chapter 6

<sup>1</sup> This point could be related to what Braddon-Mitchell/Jackson (1996) call the "depth problem" with causal theories of reference. The problem "...is to explain why 'water' picks out a property of something at a certain point, at a certain 'depth', in the causal chain that ends up with our using the

word 'water'" (p.68). Perhaps with me and my doppelganger's notional water-attitudes, the causal chain could be said to stop at the so-called "superficial" properties of our respective waters, i.e., a colourless, odourless, transparent liquid, which has a certain conceptual role within the social community.

<sup>2</sup> When I talk of the representational states of a subject, what I am referring to are internal states which represent the external world as being a certain way, at least from the subject's first-person perspective. In particular, what I have in mind is the representation of a subject's environment by their sense modalities, particularly the visual representation of the environment. I suppose that what I mean is something akin to Putnam's (1975) notion of the "stereotype" of a term such as "water". Even though it is considered that Oscar and Twin Oscar's "water" terms mean different things due to water being composed of XYZ on Twin Earth, Putnam acknowledges that their respective stereotypes of water are identical, i.e., water is a transparent, colourless, odourless, liquid etc. If one accepts the intuition about the reference of natural kind terms such as "water", the moral of the Twin Earth situation was that the Oscars' "water"-stereotypes underdetermined reference for their "water"-terms. Whether this referential under-determination of a "water" term is capable of producing some sort of feasible narrow content will be investigated in this part.

The representational content will be considered narrow, because a subject's representation of their environment, will be constructed from the skin-in, so to speak. Proximal stimulations on the subject's sensory receptors will be considered as internal states or properties of the subject. In essence, it will try to be ascertained whether a feasible narrow psychology can be constructed just from the subject's internal representations (and the way they interact with each other, so there is also a conceptual role element involved). This method of attempting to construct a narrow psychology from the proximal stimulations in, has been used before, e.g., see McDermott (1986), McGinn (1989), Devitt (1990). Even though these narrow representational strategies are interesting in their own right, I decided to concentrate on Dennett's (1982) idea of "notional worlds" to illustrate a theory of narrow representational content. Now, this choice could be viewed as questionable. For instance, Dennett himself claims that the conception of notional world psychology "...is one approach to what has come to be called "narrow conceptual role semantics"" (1987, p.209), which may lead one to ask why his notional worlds are not in the part on conceptual role narrow content, along with Loar's psychological content. I admit that this could well have happened, and that there is an element of narrow conceptual role content involved with notional worlds. However, it seems to me that Dennett's notional worlds conception is even more effective when presented as an example of a narrow representational content theory, and so this is what I have done.

<sup>3</sup> Putnam (1981) also makes the comparison between Dennett's "notional worlds" and Husserl's idea of "bracketing" on p.28. Putnam claims that with a bracketed belief "...the device of bracketing subtracts entailments from the ordinary belief locution (all the entailments that refer to the external world, or to what is external to the thinker's mind)" (1981, p.28).

<sup>4</sup> This is similar to the approach that Woodfield (1982) describes, which is one that "...seeks to specify the content of a cognitive state via the specification of conditions that the world would need to meet if the state were to represent reality truthfully... It tries to pin down content by specifying a way the world must be when the content matches it, instead of specifying what the representans itself is like" (1982, p.260).

<sup>5</sup> In examining Loar's above argument, I should point out that I am doing so purely because it involves Loar's criticism of Dennett's notional worlds as being instrumentalist in nature. Later on, I try to show how Dennett could avoid this particular criticism, which in turn leads me to suggest a different method of formulating a subject's notional world.

<sup>6</sup> This point was made by my supervisor.

<sup>7</sup> These points were made by my supervisor.

## CHAPTER 7

# A NARROW REPRESENTATIONAL REFORMULATION OF DENNETT'S NOTIONAL WORLDS

### 7.1 A Far-Fetched Thought Experiment<sup>1</sup>

Having examined Dennett's conception of notional worlds and some views/criticisms of that conception, I will now try to give a specifically narrow representational version of Dennett's idea. My motivation for doing this is that I believe the notional worlds conception is very amenable to a narrow representational reading, and may provide a very attractive and plausible conception of narrow content.

Therefore, let us imagine that I am walking down the streets of my home-town of Neath,<sup>2</sup> visiting various places to satisfy my wants and needs. At exactly the same moment on Twin-Earth, my doppelganger is also walking down the streets of his home-town of Neath (it would be Twin-Neath to us), visiting the same places and satisfying the same wants and needs as I am. With this Putnamian-type situation in place, imagine that I stop in front of Argos, and examine a rather gaudy water-feature for gardens which is in the shop window, and is having water pumped through it to show how it looks in full operation. Having gazed at it for a little while, I come to have the belief which is expressed by my utterance of "the water is coming from the

gnome's bottom". In Twin-Neath, my doppelganger is also in front of Argos and is having the same belief which is expressed by the same utterance.

At this point in thought-experiments of this kind, the following claims are usually made. My doppelganger and I are intrinsic physical duplicates, that is, we are identical with regards to neurophysiology, functional or conceptual role, and behavioural dispositions. We are also both having indistinguishable perceptual and phenomenal experiences. Indeed, the situations of Neath and Twin-Neath are almost physically identical, the only difference is that in Neath water is composed of H<sub>2</sub>O whilst in Twin-Neath water is composed of XYZ. At this point, one could say the following: despite the physical and psychological isomorphism of me and my doppelganger, with a wide propositional attitude psychology, our respective water-beliefs would be expressing different propositions. My belief-utterance would express the proposition that there is H<sub>2</sub>O coming from the gnome's bottom, whilst my doppelganger's belief-utterance would express the proposition that there is XYZ coming from the gnome's bottom. As on this view, the content of a belief is the proposition which is expressed by that belief, then the claim can be made that me and my doppelganger's respective beliefs that there is water coming from the gnome's bottom have different contents, which is due to the differences in our respective physical environments.

In relation to this point, it could also be added that my doppelganger and I's respective beliefs are different because they have different truth-conditions. As a belief's truth-condition is usually given by the propositional content of the belief, then it can be seen that if two beliefs have different truth-conditions then they would be

considered to have different belief-contents, and so be different beliefs. With me and my doppelganger's respective beliefs, it can be seen that they do have different truth-conditions. That is, my belief will be true if and only if there is H<sub>2</sub>O coming from the gnome's bottom whilst my doppelganger's belief will be true if and only if there is XYZ coming from the gnome's bottom.

I think that Dennett would agree with most of the above, indeed, perhaps he would say that the content of me and my doppelganger's respective propositional attitudes have been fixed in relation to our respective actual environments. Therefore, it can be seen that the fixing of the propositional attitudes of an agent is determined by the internal states and properties of the agent and the extrinsic relations that these states and properties have with the external world. Now, what we and Dennett want to know is what contribution just the internal states and properties of the agent make to the fixing of the agent's propositional attitudes, or put another way, how we would "characterize the narrow-psychological states" of the agent.

In order to see how this can be done, we now return to the tragic thought experiment that was previously unfolding, and which is about to take a more interesting turn of events. As me and my doppelganger are peering into our respective Argos shop-windows, unbeknownst to us there are aliens hovering in an invisible ship above us, monitoring our actions. At the precise moment that we are having our respective beliefs that "the water is coming from the gnome's bottom" the aliens freeze us, like pressing a pause button on a DVD player when a film is playing. We are frozen neurophysiologically, psychologically and experientially, from our subjective perspectives the water coming from the gnome's bottom freezes and is a perfectly still

column of water. After this the aliens transport us onto their respective ships and take us back to their home planet somewhere in the Zeta Reticuli star system.

The aliens that have kidnapped my doppelganger and I travel the universe capturing various creatures that they come across on the multitude of worlds they visit. Back on the aliens' home planet the captured creatures are taken to a giant facility which is a cross between a museum and a zoo. When my doppelganger and I reach the facility we are put into glass cabinets containing some nutrient fluid and strange devices with cables attached are put on our respective heads. The cables from the devices lead off and are attached to an alien super-computer, which has capabilities beyond human comprehension.

Therefore, the upshot is that me and my doppelganger are in a worrying *Matrix*-style situation, totally cut off from our normal environments. However, there is an important difference between the situation of the alien facility and that of the *The Matrix*.<sup>3</sup> With the latter, humans are plugged into one huge computer simulation of various places on the Earth, and their minds and virtual bodies interact with each other inside this simulation. However, at the alien facility, each of the creatures attached to the alien super-computer will be occupying their own individually simulated notional worlds. The travelling aliens can't bring all the worlds they see back to their home planet, but what they can do is bring back sentient creatures from the inhabited planets, creatures capable of a certain level of mental representation, creatures which, in essence, have representations of their worlds neurologically encoded inside their brains. If the humans attached to the Matrix are being used as batteries for producing power, then the creatures attached to the alien super-computer

are being used as organic projectors of their subjective experiences of their home-worlds.

## 7.2 A Cunning Plan for the Objective Determination of a Notional World

Having given a convenient science fiction background to the situation, we must now get back to considering the issue which this thought experiment is concerned with, namely, how is the organismic (i.e., narrow content) contribution to the fixing of a person's propositional attitudes going to be worked out. The situation that me and my doppelganger are in is essentially the same as the situation of two doppelgangers whose notional worlds are going to be formulated in Dennett's original situation. That is, all the people concerned are still alive, but are somehow frozen in suspended animation, while also being completely isolated from their respective environments. With Dennett, the person's notional world is formulated from a "Laplacean snapshot" of his internal structure and behavioural dispositions. As previously discussed, Dennett's method of notional world formulation does not depend directly on the internal subjective belief-states, etc. of the person, but is instrumentalist. In Dennett's sense, a person's notional world could be described as a possible world where the maximal number of the person's beliefs were realized; but the beliefs would only be posited and utilised indirectly, as a heuristic method for rationally explaining the person's behaviour and dispositions to behave. As stated in the previous chapter, Loar was of the opinion that a problem for this method of notional world formulation was that if there were no inner states or properties that could be said to objectively determine a person's notional world, it could fall foul of the externalist argument

described by Loar. The upshot of this would be that doubt could be cast on the claim that a person's narrow notional world was determined by internal factors.

Therefore, what options do we have for attempting to avoid the conclusion of the externalist argument? It seems to me that the first and most obvious move is to ensure that the notional beliefs and attitudes of the person whose notional world has been formulated are real internal states, and not just instrumentally ascribable entities. A related point, is that the subjective perspective of the person whose narrow notional world it is, is going to have to be utilised, in order to give the person's experiential states some semblance of intentionality, or rather quasi-representationality. If all of this is taken on board, then the remaining choice is concerned with how the person's narrow content is going to be objectively determined.

My favoured option for achieving an objective determination of the person's narrow notional world is to claim that the latter is constituted by the neurophysiological states of the person.<sup>4</sup> What I am claiming is that what is subjectively experienced as the person's notional world can be constituted from the objective perspective by the total neurophysical properties and structure of that person. I am also including in the description "neurophysical" all the proximal stimulations on the sensory receptors of the person concerned.<sup>5</sup> This means that all the neurally created electrical impulses that run from the sense organs to the brain, and whatever "content" or "information" these electrical impulses have, are being considered to be internal states. I have called the proximal stimuli and electrical signals that result from these stimuli "internal", rather than "intrinsic". The reason for this is that it might be doubted that physical states that ultimately depend for their existence on stimulations from the external



world could properly be called “intrinsic”. However, if we start the analysis from the proximal stimulations, on the “inside”, so to speak, then I think these can correctly be called “internal” and considered to be part of the organismic contribution.<sup>6</sup>

Therefore, in this thought experiment, *by hypothesis*, the aliens and their super-computer are basically able to construct me and my twin’s subjective notional worlds from our objective neurophysical properties and structure. If this is accepted, it enables one to say that because me and my twin are neurophysically isomorphic, then the quasi-representational content of our respective subjective experiences are also isomorphic. In making this claim, I am taking the “quasi-representational content” of a person’s particular mental state to be determined by how things seem to that person, or how that person conceives things as being. Moreover, the person’s mental content only has *quasi-representationality*, as I take genuine semantic representationality to involve the mental content having a causal-referential relation to the object or state of affairs in the external world that is being represented.<sup>7</sup>

This point leads on to a related issue. That is, even though me and my doppelganger’s neurophysiology is identical, and from our subjective viewpoints the contents of our quasi-representational states are also identical, I am not denying that on a referential truth-conditional reading of our respective representational states, it could be claimed that they did in fact have different contents (I will come back to this point in a short while). As before with the subject’s proximal stimulations, it could be questioned whether the subject’s quasi-representational states could be considered “internal” in the sense required for situations like this. The usual line is that the subject’s internal states must be “non-intentionally described”, or as McGinn (1989) puts it, the internal

facts "...must not...be specified by reference to the subject's environment" (McGinn, 1989, p.2).

However, as with the proximal stimulations, if we consider the quasi-representational states from the inside or from the skin in, so to speak, then I think that they can correctly be called internal states. After all, the reason that I am prefixing "representational states" with the word "quasi-" is that the states in question are not fully intentional or fully representational; that is, they are not being considered to be referring to objects in the subject's actual environment. Considering quasi-representational states as internal, means that they are considered in isolation from the person's environment, they do not have the full semantic content of normal representational states. Therefore, in what follows, I will mention, for example, the gnome in me and my twin's notional worlds, which seems to be a case of our internal states being specified by reference to our respective environments, which is contra McGinn above. However, as far as I am concerned, that is still going to be an internal state, as what I will be referring to will be the proximally stimulated gnome-experience on my visual receptors.

### **7.3 Notional World Formulation Using the Cunning Plan**

Having given my preferred option of how to formulate a person's notional world, it is now time to consider the actual formulation of me and my doppelganger's respective notional worlds. Our notional worlds will be formulated around what I will call the "Argos-episode", where my doppelganger and I have been frozen as we are staring

into our respective Argos shop-windows and are having our respective beliefs that “the water is coming from the gnome’s bottom”. To achieve this we can imagine the following. My doppelganger and I have been put into glass cabinets next to each other, as the aliens have intriguingly noticed that we both bear a striking similarity of appearance to each other. The alien super-computer that we are both attached to is then activated and our respective notional worlds are formulated. With Dennett, a creature’s notional world is constructed on paper, but with the alien super-computer it is formulated as a 3-D projection/simulation from me and my doppelganger’s neurophysiology (which of course, includes the proximal stimulations). Therefore, what is happening is that what we are experiencing from our respective subjective viewpoints at a particular time, is being translated from our respective neurophysiologies and objectively simulated, so that others can view our subjective experiences.<sup>8</sup>

Despite the strange alien environment that my doppelganger and I are located in, it seems plausible to me that the notional worlds that we will subjectively experience, and which will be objectively simulated for the aliens to observe, will be that of the respective Neath and Twin-Neath Argos shop-windows and gaudy water-features. Our respective notional beliefs about the water coming from the gnome’s bottom will also be reactivated, and the notional objects of these beliefs will be the respective gnomes on the water-feature.

In relation to the above conclusion, it will be recalled that Stalnaker criticised Dennett’s conception of notional worlds, formulated only on “...an internal description of the movements” of a person. In particular, he asked how we could get

any information about Boston without referring to things outside the person's skin. However, as I hope I have shown, if the proximal stimulations and quasi-representational states of the person are allowed to be called internal states, then the organismic formulation of a person's notional world could claim to involve things that are outside the person's skin. With me and my doppelganger's formulated notional worlds, the description of those worlds are containing some information about Neath and Twin-Neath respectively, even though it could be claimed that only internal states have been used in their formulation. Perhaps we couldn't say that actual denotation of external things was going on; but we could say that if the person's experiences are assumed to be veridical, then given the person's proximal stimulations and quasi-representational states, the resulting notional world that is formulated will be the most plausible environment outside the person's skin. Of course, at the alien facility my doppelganger and I's experiences are not veridical, but I don't think this point counts against the previous claim. For instance, if we were both back in Neath and Twin-Neath, outside our respective Argos stores, then our proximal stimulations and quasi-representational states would create notional worlds that were good indicators of what our environments were like.

Stalnaker makes another related criticism, where he claims that the movements a person is disposed to make that may seem appropriate for facts about Boston, are instead appropriate for completely different facts, e.g., "...facts about the social organization of termite colonies" (Stalnaker, 1989, p.185). The point that Stalnaker is making here is that the content of our internal states is normally constrained by facts in our actual environment (Stalnaker, 1989, p.184). Put another way, it could be said that the ordinary wide content of our internal mental states constrains what content

those states could be said to have. It can be seen that the direction of content-constraint here is from the outside to the inside, i.e., the actual environment determines what content a person's intentional states have (the latter is basically the essential claim of the externalists). However, with Dennett's notional worlds, the direction of content-constraint is in the opposite direction, i.e., from the inside to the outside. It is from organismic internal properties, like behavioural dispositions, that the notional world of the person is formulated. It is this process that Stalnaker thinks is "too unconstrained to work", as there is no involvement of the person's actual environment in the formulation of his notional world. This enables Stalnaker to make the above claim, that even though the person's narrow internal dispositions may seem appropriate to the facts of Boston, the actual wide content of those narrow dispositions could be something completely different.

Silverberg (1995) discusses this particular criticism of Stalnaker's, where he claims that the notional or narrow thought-content shared between two internal physical duplicates would mean that things seemed the same to each duplicate, i.e., how their notional worlds seemed to each of them would be the same for both of them. This would be the case even if their respective wide thought-contents were different, e.g., one duplicate could have Boston as his wide content, whilst the other could have a termite colony as his wide content (Silverberg, 1995, p.118). It seems that Silverberg quite happily acknowledges that the wide thought-content of two physical duplicates could be completely different, yet the important point to him appears to be that as far as the duplicates are concerned, the narrow content of those same thoughts does not seem to the duplicates to have anything to do with termites and their colonies (Silverberg, 1995, p.118). I suppose that this is a fair point, however, I don't think

that it really deals with the point that Stalnaker is making about narrow content being too unconstrained to be viable. After all, Stalnaker's claim is that despite the person's narrow notional world seeming a certain way to him, the actual situation is that his actions are really appropriate to the facts concerning a different wide content, that of a termite colony.

It seems to me that with Dennett's original method of formulating a person's notional world, it is going to be difficult to counter this particular criticism of Stalnaker's.

Dennett's thought experiment involved the notional world theorists as just receiving the person (or creature) which had been frozen, not knowing what its actual environment had been like. This point seems to be the crux of Stalnaker's criticism that Dennett's narrow notional worlds are too unconstrained to work. However, with my slightly reformulated version of Dennett's notional worlds, there might be a way to escape Stalnaker's criticism. One difference is that with my thought experiment the person's actual environment is made known, the person is not just served up to the aliens without us knowing what his environment had been like. The person is frozen in his actual environment, just at the moment that he is having certain beliefs and experiences. With me in Neath, the particular moment is when I am standing in front of Argos's shop window looking at a gaudy water-feature, and I am having the belief that the water is coming from the gnome's bottom. Therefore, at the moment that I am frozen by the aliens, we all know what the wide content is of my internal psychological and neurophysiological states.<sup>9</sup> Thus, contra Dennett's thought experiment, any narrow content that my internal states may have, has already been fixed by my actual environment.

In admitting this, I don't think that I am creating a problem for the consequent formulation of my notional world. After all, Dennett agrees with Stalnaker that the content of our ordinary propositional attitudes is wide; what he is trying to find out is what the organismic (narrow content) contribution is to the fixing of our propositional attitudes. In the Argos-episode, I am frozen and removed from my actual environment in the process of having a wide propositional attitude. What the aliens will discover is what contribution my internal physical states (in which are included the proximal stimulations and quasi-representational states) are making to the having of that wide propositional attitude. Whatever contribution this is, characterises my narrow psychological states. However, in relation to Stalnaker's particular criticism, the upshot of this is that my actual environment is included in the thought experiment; indeed, we can suppose that it is causally responsible for my proximal stimulations and internal representational states. If this is accepted, then it can be claimed that the consequent narrow notional world that is formulated using these proximal stimulations and internal (quasi-)representational states will have an apparent wide content that is constrained by the actual environment I had been embedded in.<sup>10</sup>

At this point, one of Dennett's most important ideas must be taken on board. He makes the point that in order to do psychology one must have some knowledge of the "*semantic* properties of the internal events and structures under examination" (Dennett, 1982, p.40), which will involve "looking at the relations of those internal events or structures to things in the subject's environment" (Dennett, 1982, p.40).<sup>11</sup> In our ordinary, everyday situation, this is accomplished by matching up our internal mental representations with their objects in the actual world. However, this is not

possible in the artificial environment that my doppelganger and I are presently located in. But Dennett has a solution:

...nowhere is it written that the environment relative to which we fix such a system's semantic properties must be a *real* environment, or the *actual* environment in which the system has grown up. A fictional environment, an idealized or imaginary environment, might do as well (Dennett, 1982, p.40).

Or possibly a super-computer simulated environment? It seems that we have exactly what Dennett says we need. Not only that, but it has also been formulated according to Dennett's strictures. That is, the notional world episodes have been formulated using our internal representations, but it is not the representations that my doppelganger and I are experiencing in our notional episodes, but rather what would be considered to be the notional objects of those representations, given that those representations are a certain way (Dennett, 1982, p.40).

So in the notional Argos-episode, my quasi-representation of the gaudy water-feature would have a simulated gaudy water-feature as the notional object or referent of that quasi-representation. From my subjective perspective, there would be the appearance of intentionality or representationality,<sup>12</sup> with the accompanying appearance of "real" semantic properties. With this situation, it can be seen that a person's notional environment will depend on the content of that person's internal representations.<sup>13</sup> This is in contrast to the ordinary situation, where it could be claimed that the content of a person's internal representations will depend on that person's environment. The situations are exactly converse because we are using what Dennett calls the "proto-semantics" of a notional world to try to work out what the organismic contribution is (Dennett, 1982, p.40).



#### 7.4 Specifying the Organismic (Narrow Content) Contribution to a Person's Propositional Attitudes

It is now time to try to answer Dennett's original question, i.e., what is the organismic contribution to a person's propositional attitudes? The answer to this question will be equivalent to the answer of what narrow representational content is shared between my twin and I in our respective notional Argos-episodes. I am only considering notional episodes for the sake of simplicity; if we can figure out what narrow content is shared between the notional episodes, then perhaps we could just expand this point to what narrow content should be shared generally between me and my twin's notional worlds. If what has gone before is accepted, then it seems plausible to claim that both our formulated and projected notional Argos-episodes will have qualitatively identical appearances; not only from the subjective viewpoints of my twin and I, but also to the objective viewpoint of the aliens who are observing our projected notional worlds. Therefore, what will be experienced and observed are the respective Argos shop-windows with the gaudy water-features in them. My twin and I will also be entertaining our respective beliefs that the water is coming from the gnome's bottom. Considering that to achieve all of this, we have only used internal neurophysical states considered in isolation from the actual environment, it seems that the organismic contribution to our propositional attitudes is quite considerable.

However, having said this, all I get in my notional episode are the surface appearances of objects and states of affairs. My experience is perhaps of a representational superimposition that is not actually superimposed over the external objects and states of affairs that it is supposedly representing. If I was embedded in my actual

environment, then the representational superimposition that I was experiencing would be superimposed over the external objects and states of affairs that it was supposedly representing.<sup>14</sup> The problem that this situation raises for the reference and meaning of terms and thoughts, occurs if the identity conditions of the concept or entity that is being represented, depends on more than just its surface appearance. With natural kind concepts, for example, it is accepted that their identity conditions depend on the internal microstructure of the natural kinds, i.e., what sort of stuff they are actually composed of. Because water is considered to be a natural kind, as there was a difference in the internal microstructures of Neath water and Twin-Neath water, that was the reason that my twin and I's respective propositional attitudes were earlier considered different. So there were differences between our propositional attitudes, differences of propositional belief content, even though our perceptual and phenomenal experiences were qualitatively identical.<sup>15</sup> The upshot of this is that if all we have to determine meaning/reference are the surface appearances of objects, then on some occasions the meaning/reference of some of our terms and thoughts would be underdetermined, we would be incapable of having "ultimate reference".

Therefore, the problem is how to specify what narrow content is shared between me and my twin's respective notional attitudes, the latter of which are dependent only on surface appearances for the determination of their meaning/reference. What should also be borne in mind is that whatever reading we give to our notional attitudes, that reading will still have to be possible if me and my twin were embedded in our original environments. For that reading will essentially be the narrow organismic contribution to the fixing of our wide propositional attitudes. The situation we want is that even though my twin and I are in different environments and so our wide propositional

attitudes concerning “water” will be different, our narrow notional attitudes concerning “water” will be the same.

#### 7.4.1 *Can Notional Beliefs be Given a De Re Specification?*

Therefore, what reading can be given to my twin and I’s respective notional beliefs that can be expressed by the predicate “believes that water is coming from the gnome’s bottom”. I don’t think that the notional “water”-beliefs can be given a de re reading, where the general term “water” has the function of specifying, or referring to, the particular stuff called “water”; at least if we are trying to discover a shared content between the respective notional beliefs. The reason for this is that a de re general term in a belief predicate, will be individuated by the object-type or substance-type that the term is referring to. In other words, the term will be individuated by its wide content. Therefore, my notional de re “water”-belief will be individuated by H<sub>2</sub>O, while my twin’s notional de re “water”-belief will be individuated by XYZ.

Another related point that seems to show that a de re reading of our respective notional “water”-beliefs will not capture a “content” that is shared between them is that on a de re reading, the term “water” is considered to have a transparent or non-oblique occurrence. This means that a co-referential term for “water” could be substituted in our notional “water”-beliefs without changing the truth value of the containing sentences. For instance, on a de re reading, this would mean that it is true that I notionally believed that water was coming from the gnome’s bottom, whilst it would also be true that I notionally believed that H<sub>2</sub>O was coming from the gnome’s

bottom (and the same for my twin, except there it would be XYZ). However, this might not be the case, especially if my twin and I had no idea of the chemical compositions of our respective waters.

Therefore, it can be seen that a *de re* belief ascription is not the right one for trying to capture our common subjective conceptions of things contained in our notional “water”-beliefs. Indeed, *de re* belief ascriptions are more appropriate for getting information about the subject’s world, rather than characterising the subject’s conception of the world. What we want is a characterisation of our notional “water”-beliefs that does not discriminate between Neath and Twin-Neath, i.e., the water the notional beliefs are about could either be H<sub>2</sub>O or XYZ.

#### 7.4.2 *Can Notional Beliefs be Given a De Dicto Specification?*

What about a *de dicto* reading of my twin and I’s respective notional “water”-beliefs? At first sight, this strategy sounds more promising, as a term understood in a *de dicto* manner does not always directly specify, or refer to, the object or substance that it is about. A *de dicto* reading of a term is usually meant to characterise how the subject conceives things as being when using the term, i.e., it is meant to capture how the subject’s belief represents the object that it is about, to the subject himself. Because of this, a *de dicto* reading of a term in a belief usually means that the term has an opaque or oblique occurrence in the containing sentence. This means that when the term in question is substituted for a co-referential term, then the truth value of the containing sentence may not stay the same. For example, if I do not know the

chemical composition of water, then even though it may be true that I believe that water is coming from the gnome's bottom, it could be false that I believe that H<sub>2</sub>O is coming from the gnome's bottom (and the same for my twin, except there it would be XYZ). This sounds just right, on a de dicto reading of our notional "water"-beliefs, my twin and I could be believing something in common about "water", as our respective "water" terms are not inter-substitutable, in a truth-preserving way, with their respective co-referential terms. After all, it was the extensions of our "water" terms that were different, so perhaps we could now say that the wide content of our notional "water"-beliefs is captured on a de re reading, whilst the narrow content of our notional "water"-beliefs is captured on a de dicto reading.<sup>16</sup>

However, even this option seems closed to us, at least if we accept the arguments of Burge (1979, 1982), who argues that even on de dicto readings, me and my twin's notional "water"-beliefs would have distinct contents. The reason for this, if I have understood Burge correctly, seems to be the following. The crux of it is the definition that Burge gives to an attitude content, which he takes as "...the semantical value associated with oblique occurrences of expressions in attributions of propositional attitudes... Thus, the content is, roughly speaking, the conceptual aspect of what a person believes or thinks" (Burge, 1982, p.119). Burge then argues that while I could have a de dicto "water"-belief ascription with a particular conceptual content for my term "water", my twin's de dicto "water"-belief ascription would have a different conceptual content for his term "water", and so one could not use the term "water" in oblique position in my twin's "water"-belief ascription, one would have to use *twater* or *twin-water*, etc (Burge, 1982, pp.109-110). Because our respective waters are chemically composed of different stuff, then the concepts associated with

them, and so the semantical value attached to the oblique occurrences of “water” in our respective de dicto “water”-belief ascriptions would be different. Hence, the upshot of this is that on a de dicto reading, my twin and I’s respective notional “water”-beliefs would still have a different content, at least with regards to our “water” terms.

What is needed is a narrow semantical value that can be shared between my twin and I’s respective notional “water” terms, a value which basically can only be taken from the surface appearances and proto-semantic properties of water that are allowed to us from our respective notional Argos-episodes. What we are trying to do here is to internalise or bring back into the skin the referential and semantical properties which we associate with our respective waters. The externalist method of providing the reference and meaning for a natural kind term will result in wide content. For example, with a natural kind term such as “water”, the referent of that term will be said by the externalist to be a particular object or substance in the subject’s actual environment that has a certain microstructure. The externalist would also claim that what we “mean” by our term “water” will also be an object or substance that has a certain microstructure. Of course, the externalist could also acknowledge that the substance which was referred to with the term “water” also had certain so-called “superficial” properties of appearance, but that these were not essential to the identity conditions of water. Now, I am not denying any of the foregoing, indeed, it was by intuitively plausible reasoning (and thought experiments) such as this in the work of Kripke and Putnam, that the causal theory of reference was seen as preferable to the description theory of reference (at least as far as the reference of natural kind terms were concerned).

It is interesting to note that if we ignore the “water” terms in my twin and I’s notional and propositional attitudes that water is coming from the gnome’s bottom, then the narrow organismic contribution to our respective propositional attitudes is not inconsequential. Indeed, the organismic contribution seems to provide most of the proposition believed, except for the “water” part. Therefore, it could be claimed that just the contribution of the narrow belief content produces our respective propositional attitudes that “(something) is coming from the gnome’s bottom”. This seems like quite a result, although it does sound fairly disgusting. However, it seems to me that if we want to attain a narrow organismic sense of the reference and meaning of my twin and I’s respective “water” terms, so that we can both properly use the term “water” in our notional attitudes, then we will have to return to something like the description theory of reference for our respective notional waters.

This basically means going back to something like Putnam’s (1975) “stereotype” of water, which is that water is a transparent, odourless, tasteless, colourless liquid, which flows from taps, is drinkable, we wash in it, etc. If this strategy is adopted, then it allows us to claim the following. When my twin and I’s “water” terms are used in our respective notional belief predicates, the reference of those “water” terms can be said to be the liquid in our respective environments which satisfies the stereotype or definite description for water.<sup>17</sup> The advantage of this will be that instead of “water” having a referent which is a particular liquid with a particular microstructure, we are now referring to whatever liquid satisfies our respectively identical stereotypes for water. This of course means that the referents for me and my twin’s “water” terms are being underdetermined, but this is what we want with our

notional “water” terms, as they should not discriminate between water in Neath and water in Twin-Neath. We can let our different wide propositional attitudes refer correctly to water in our environments, but the narrow notional attitudes can be what is shared between us and what makes up the organismic or narrow content contribution to our respective propositional attitudes.

In a similar vein, we can also say that the meanings of our respective “water” terms are not fixed to semantical concepts which essentially contain reference to the particular microstructure of the liquid in question. The semantical values associated with our respective “water” terms could use the stereotype for “water” (and possibly the use or conceptual role that water has in our lives).<sup>18</sup> For example, both my twin and I could say that what we mean by “water” is that transparent liquid in our environment, which we drink and wash in, which falls from clouds, and fills oceans and rivers, etc.<sup>19</sup> The upshot of all this is that my twin and I could be allowed to have de dicto notional beliefs containing the term “water” in oblique position, as the internalist versions of meaning and reference which we have used for those “water” terms, means that they have the same conceptual and semantic values. Moreover, it could also be claimed that all this has been achieved only from the surface appearances and proto-semantic properties that have been allowed from my twin and I’s organismically formulated notional Argos-episodes.

Therefore, the situation we have is the following. In our respective notional Argos-episodes, my twin and I share a narrow content that can be summed up by the notional attitude “believes that water is coming from the gnome’s bottom”. Moreover, our respective notional attitudes could be viewed as being de dicto and as having the term



“water” in oblique position, which would mean that (according to Burge) this would provide and characterise the contents of our mental states (Burge, 1979, p.76, p.87; 1982, p.99). Indeed, perhaps we could say that the shared notional water attitudes were analogous to the subjective de dicto beliefs that I earlier argued were shared between actual-Bert and counterfactual-Bert in Burge’s “arthritis” thought experiment.

If all this is accepted, then it makes one wonder why Burge would claim that two twins could not have the same de dicto water-thoughts with “water” in oblique position, due to differences in the chemical composition of the respective waters in their different environments (Burge, 1982, pp.108-111).<sup>20</sup> What I am getting at is that if obliquely occurring terms in de dicto beliefs are acknowledged to have the function of showing how the subject is conceiving of things, or of how the subject is representing things, then how can one ascribe beliefs to the subject which have properties that the subject himself is unaware of. In making this claim, I realise and accept that my twin and I’s wide propositional attitudes will have conceptual and semantic properties that we are not aware of. That is, the “water” in my water-belief will have a semantical value of H<sub>2</sub>O, whilst the “water” in my twin’s water-belief will have a semantical value of XYZ.

What I am questioning is how an oblique occurrence of “water” in our (possibly) de dicto notional attitudes can be claimed to be utilising H<sub>2</sub>O and XYZ respectively, in providing and characterising the conceptual/semantic content of our “water” terms from the *subjective viewpoints of me and my twin*, as we don’t know what the microstructures of our waters are (At least, this is what I think Burge is claiming in

his 1982). For example, my notional attitude involves using “water” obliquely, which means that one could not substitute a co-extensive expression (H<sub>2</sub>O) for it *salva veritate*. My twin’s notional attitude is also using “water” obliquely, which means that one could not substitute a co-extensive expression (XYZ) for it *salva veritate*. What this means is that me and my twin would assent to holding the respective beliefs that water is coming from the gnome’s bottom, but we may not assent to the respective beliefs that H<sub>2</sub>O/XYZ was coming from the gnome’s bottom (especially if we did not know any chemistry).

To me, this seems to show that the microstructures of our respective waters are playing no role in providing the conceptual or semantic characterisation of me and my twin’s notional water-beliefs, at least from our subjective perspectives.<sup>21</sup> Indeed, if the only differences between our waters are not involved in the subjective characterisation or representation of our notional water-beliefs, then it seems plausible to claim that the conceptual and semantic properties for those notional beliefs are coming from the respective water-stereotypes which we have, i.e., water is transparent, colourless, etc.<sup>22</sup> Perhaps there is such a thing as a subjective *de dicto* belief which genuinely characterises how the subject is conceiving things. This would be in contrast to an objective *de dicto* belief *ascription*, which would characterise the expressions in the subject’s belief according to their conventional meanings (this is again similar to Loar’s distinction of psychological content and social content in *de dicto* beliefs).

## Objections to the Stereotype-Strategy of Specifying Narrow Content

However, it must be admitted that there are objections in the literature to the above strategy of using something like the stereotype of water to constitute the narrow meaning and content of “water”-beliefs. For example, McCulloch (1995) makes the amazing claim that:

...post-Putnam it is no longer permitted to rewrite (1) and (2) like this [this rewrite of (1) and (2) are belief ascriptions using the stereotype, or what McCulloch calls the “Lockean definition”, of water] in order to display the shared belief posited by the internalist. For Putnam’s key point is that there is no such shared definition or description which will adequately convey what is understood by ‘water’ (McCulloch, 1995, p.197).

This news is going to come as quite a shock to the billions of people in the world who have daily contact with water and who would plausibly give the meaning of “water” as the clear, transparent, liquid that comes out of taps, that we drink when thirsty, that we wash in, etc. This seems to be a case of philosophical theory gone mad, where the philosophical theory is Putnam’s (and Kripke’s) theory on the individuation of natural kinds and the use of natural kind words. McCulloch’s claim of impermissibility in using the stereotype of water to provide the narrow meaning of water beliefs only goes through if it is accepted that “water” means H<sub>2</sub>O. It is this latter claim that I have been arguing against in the previous paragraphs. I am not denying that water is chemically composed of H<sub>2</sub>O, all I am denying is that the meaning of “water” is exhausted by its chemical composition.

Putnam’s (1975) claim was of course that the stereotype of water did not provide necessary and sufficient conditions for giving the meaning of water qua H<sub>2</sub>O. For

example, on Twin Earth, H<sub>2</sub>O may have the “superficial” properties of being a black, tarry, substance, so Putnam would claim that the intuition here is that this is what water would be on Twin Earth. Therefore, the Earth stereotype of water is not necessary for something to be water. Meanwhile, the existence of XYZ on Twin Earth, which has identical “superficial” properties to Earth water, is not considered to be water by Putnam. Therefore, the Earth stereotype of water is not sufficient for something to be water.

However, as I have said, this only follows if we go along with the intuitions that Putnam (and Kripke) want us to have about the meaning of the term “water”. The intuition that I get is that in everyday life, the so-called “superficial” properties associated with the stereotype of water are integral to giving the meaning of “water”. With regards to H<sub>2</sub>O being a black, tarry, substance, I personally would not consider that substance to be water, whilst with XYZ, the intuition I get is that this could be a different type of water.

Putnam’s essentialist view of natural kind individuation and meaning came from Kripke’s arguments about the reference of proper names, which has been mentioned previously. It will be recalled that Kripke argued that the proper name “Plato” would still refer to the individual who was originally “dubbed” as Plato, even though someone else might satisfy the definite description that is associated with Plato, such as being the pupil of Socrates and the teacher of Aristotle. It seems to me that Kripke’s intuition is strong here, as we would not be tempted to describe the individual who has the “superficial” properties associated with Plato as a different type of Plato, so to speak.

However, I don't think the intuition is as strong in relation to natural kinds (e.g. water) and the words used to refer to them. It seems far more plausible to believe that if a different substance, say XYZ, has the same so-called "superficial" properties as the substance H<sub>2</sub>O, then it might be correct to use the term "water" to refer to both substances. Therefore, the meaning of "water" would not track the essences of the substances in question, but would instead track their stereotypical properties. This result seems more intuitively plausible due to the fact that water has such an abundance of perceptual and conceptual connections to humans and their daily lives. Moreover, it is because the stereotype of "water" does not track the essences of H<sub>2</sub>O and XYZ, that it can be used to specify the narrow meaning that is shared between "water"-beliefs on Earth and Twin Earth. So, while it might not be permissible for McCulloch to give the stereotype of water as its (narrow) meaning, I think I'll join the majority of other people on the Earth who seem to have no problem in doing this.

Another objection to using the stereotype of water to specify the narrow content of "water"-beliefs is set forth in Lepore/Loewer (1986), where they state:

The suggestion was that the solipsistic interpretation of "Water is wet" is that the liquid that people drink, fills oceans, and so forth, is wet. It is clear that the content of this stereotype is not sufficiently narrow to be solipsistic. The expressions "oceans" and "people" have different meanings for Arabella and Twin Arabella. For Arabella, "people" refers to Earthlings, whereas, for Twin Arabella it refers to Twin Earthlings. The same point applies to "oceans" and "liquid" and perhaps to other concepts in the stereotype. By imagining suitable differences between Earth and Twin Earth, while keeping constant the ways things seem to the twins, it looks as though Twin Earth arguments will succeed in showing that no natural kind term has SS [solipsistic semantics] (Lepore/Loewer, 1986, p.606).

By the expression “solipsistic semantics” Lepore/Loewer mean a semantics that when “...assigning meanings to representations it does not presuppose the existence of any mental or physical individuals other than the thinker and his thoughts” (Lepore/Loewer, 1986, pp.595-596), or more concisely, “...that the determinants of the meanings of one’s mental representations are entirely within oneself” (Lepore/Loewer, 1986, p.597). Now, if we stick to the above definitions for a solipsistic semantics, then it does seem correct to say that the stereotypical meaning of “water” is non-solipsistic, as the concepts involved have meanings that are dependent on objects and states of affairs outside the subject himself. But does that mean there could not be a narrow content that is defined as being a content that is shared between doppelgangers in the standard Twin Earth situation? What I am trying to say here is that even though the content of me and my twin’s “water” concepts may be externalist or anti-individualist in the sense that they depend on our respective environments, could it not still be claimed that there is a shared narrow content that our “water” concepts have, as narrow content is defined as being a content that is shared between doppelgangers?

With the above objection, Lepore/Loewer have in mind what they call a “phenomenological strategy” of constructing a narrow content, where “...how things seem or how they appear” (Lepore/Loewer, 1986, p.606) to doppelgangers are identical. As words like “seem” and “appear” are used, I will take it that one could also say that the phenomenological strategy that Lepore/Loewer have in mind could also be described as one where how things are *conceived* as being, is the same for two doppelgangers. Therefore, let us now examine the objection that is raised against using narrow stereotypes in the above quotation. The crux of the objection is that the

expressions used in a stereotype for a natural kind term may well have different meanings for the individuals involved in Twin Earth situations. So, as Lepore/Loewer claim above, a stereotype for “water” consisting of something like “the liquid that people drink, fills oceans, etc.” is not narrow due to expressions like “people” having a different meaning for the Twin Earth doppelganger, i.e., on Earth, the expression “people” refers to Earthlings, whilst on Twin Earth, the expression “people” refers to Twin Earthlings. Lepore/Loewer also mention the expression “oceans” but do not expand on this; however, I think we can take it that they probably mean that on Earth the expression “oceans” refers to large expanses of water (H<sub>2</sub>O), whilst on Twin Earth the expression “oceans” refers to large expanses of Twin water (XYZ).

It seems to me that the first thing to say about the above is that Lepore/Loewer seem to be begging the question against the narrow content theorist. For example, to claim that “people” differs in meaning for say, me and my twin, because for my twin the expression refers to Twin Earthlings is to blatantly accept an externalist view on content and meaning.<sup>23</sup> A narrow content theorist need not disagree that the wide content of the expression “people” is different for me and my twin, but this does not entail that there can’t also be a narrow reading of the expression “people” (and the same goes for an expression like “oceans”). For instance, the narrow content theorist could say that both the Earth and Twin Earth expression “people” refer to sentient creatures that typically have certain characteristics which might include having heads, arms, legs, they wear clothes, use their mobile phones, have names, and so on. Now, I realise that the narrow stereotype for “people” I have just given is a bit rough and ready, but you get the basic idea of how it would go. However, what would Lepore/Loewer’s response to it be? Judging from what they say in the above quote it

seems that they would probably claim that the expressions used in the “people” stereotype could once more have different meanings on Twin Earth, and so on ad infinitum, for even the narrowest possible stereotype that could be provided.

But does the narrow content theorist have to accept this sort of argument? It seems to me that what is being put forward here is quite frankly nonsensical. By hypothesis, how things seem to me and my twin, that is, how we conceive of things, is supposed to be identical. Therefore, Lepore/Loewer can’t claim that the “people” on Twin Earth don’t have heads, or are green and scaly, etc. But what if they make a claim that, say, on Twin Earth the expression “head” doesn’t mean what it does on Earth, but refers to a creature that we would call a squirrel on the Earth. However, it seems to me that this claim does not affect a narrow construal of the meaning of the expression “head”. It might well be able to be claimed that the wide meaning of “head” on Twin Earth is now different to what it is on Earth; but by hypothesis, me and my twin’s conceptually and phenomenologically derived content/meaning of “head” is going to be identical. After all, if how me and my twin conceive things were different, then we would not be conceptual duplicates, which I presume we are supposed to be. I think this will be the case in any of the standard Twin Earth scenarios that involve neurophysical and conceptual doppelgangers. For instance, consider me and my twin’s shared narrow belief that water is coming from the gnome’s bottom. It wouldn’t matter if gnomes on Twin Earth were actually living creatures, and that the particular gnome my twin was looking at was called Kevin and was doing a degree in media studies, we would still have identical appearances and conceptions connected with the term “gnome”. That is, we would both conceive our



respective situations as involving a certain object called a “gnome” which had water coming from its bottom.

Of course, it might be claimed here that what is happening with the above is that the *narrow content* between my twin and I is being kept the same. However, I would deny this, and claim that it is only the conceptual properties of my twin and I which are kept identical (as well as the sensory and phenomenological properties), which is usually the case in standard Twin Earth thought experiments. I suppose that the crucial assumption with the above is that I am taking it that neurophysical duplicates will also be conceptual duplicates, or put another way, that the conceptual properties are supervenient on the neurophysical properties. The upshot of all this is that it seems to me quite feasible to use something like the Putnamian stereotype of a natural kind term like “water” to specify some sort of narrow content. That is, even though the meaning of certain terms like “water” are not entirely solipsistic, contra Lepore/Loewer, there can still be specifications of narrow content relating to that term, as that narrow content is defined as being a content that is shared between doppelgangers.

Therefore, in this chapter I have given a reformulated version of Dennett’s notional worlds, which essentially utilises the first-person perspective of the subject in formulating his/her notional world. I also attempted to give a specification of what shared narrow content would be contained in two duplicates’ identical notional worlds.

## 7.5 PART 3 CONCLUSION

This part has examined Dennett's idea of notional worlds, which I have taken to be an example of a theory of narrow representational content. The notional worlds conception can also be viewed as being part of a two-factor theory of content. The subject's notional attitudes are to be construed as the narrow component, whilst the subject's propositional attitudes, which the former are meant to determine, are to be construed as the wide component. Dennett's strategy of only using the subject's internal physical states and behavioural traits to formulate the subject's notional world seemed to me a very intuitive one. It seemed plausible to believe that the result of such a formulation would provide an answer to the question of what the organismic or narrow content contribution is to the subject's propositional attitudes.

However, the method that Dennett used to carry out this notional world formulation seemed problematic, as it appeared that the subject's notional attitudes, qua internal and causally efficacious states, would not play a role in the formulation of that notional world. Indeed, the attribution of a subject's notional attitudes is instrumentalist, i.e., the notional attitudes were being attributed mainly based on what the subject's behavioural dispositions were, perhaps as a heuristic element to help explain behaviour, but would not be deemed to have an ontologically objective existence. Perhaps this view is unfair to Dennett, with a more subtle reading possibly claiming that there really are subjective notional attitudes, but that they are constrained by objective factors. Nevertheless, it seems pretty clear that his strategy of hetero-phenomenology involved staying clear of subjective first-person sensory experience, at least in the process of formulating the subject's notional world.

It was at this point that I thought that Dennett's notional worlds could have a specifically narrow representationalist construal. It seemed to me that a plausible strategy of formulating a subject's notional world could utilise his subjective first-person notional attitudes, as long as these attitudes were constrained by a third-person objective factor, in this case the subject's internal neurophysical states (thus trying to keep to the spirit of Dennett's original idea). In doing this, I made the arguable decision to include the subject's internal or quasi-representational states in the inventory of internal properties that could be used to formulate his notional world. That is, the states were quasi-representational because they were being considered to originate from the proximal stimulations on the sensory receptors of the subject concerned. This is an arguable move, because these states seem to be referring to objects and states of affairs beyond the body of the subject, which seems to violate the restrictions on what can properly be called internal states in these situations. I based this move on the claim that the quasi-representational content of the notional attitudes was not essentially or logically dependent on objects and states of affairs in the external world. I suppose that this could be construed as being an internalist presupposition on my part; however, to claim the converse, seems to me to be presupposing externalism with regards to mental content.

The next problem was then how to specify what the narrow notional content was that was shared between me and my twin. I came to the conclusion that something similar to the subjective *de dicto* belief ascription could be utilised here. With this ascription, what our respective "water" terms would have in common would basically be the stereotype of water, which would include the so-called superficial properties of our respective waters. The properties that would individuate our narrow water-beliefs

would be ones like the clear, odourless, drinkable, liquid in our respective environments, and not the different microphysical properties of those waters. The upshot of this was that the narrow representational properties used to construct our notional worlds did seem to make a sizeable contribution to determining the wide propositional attitudes that we had about our respective waters, i.e., the organismic contribution to a person's propositional attitudes was significant.

Therefore, we now come to the question of whether Dennett's notional worlds satisfy all the conditions of adequacy for narrow content. Regarding Dennett's original conception, it seems fair to say that conditions (2) and (3) are satisfied. A plausible specification of the narrow content shared between duplicates can be given and this can be used in a psychological generalisation that subsumes the behaviour of the duplicates. However, I don't think that the original conception satisfies condition (1), that is, the narrow content is not an internal or "in the head" content that both duplicates share. The reason for this is of course that Dennett used an instrumentalist method for attributing beliefs to the person whose notional world was being formulated. Now, my reformulated version of Dennett's idea was an effort to solve this problem, but as it originally stands, his idea of notional worlds is not an adequate narrow content.

## Notes to Chapter 7

<sup>1</sup> I light-heartedly call the following thought experiment "far-fetched", and I suppose that it is a little far-fetched, but I don't think this counts against it. In contrast, Margolis/Laurence (2002), while criticising Lewis' (1994) case for some form of narrow content, claim that "After all, twins, brains in vats and swamp creatures are all bizarre philosophical inventions. Do we really want our theorizing about the mind to be dominated by them? Perhaps the right thing to say is that our intuitions in these cases are not of much interest and that they may simply have to be disregarded" (2002, p.61). When I read this, I couldn't believe it, are Margolis/Laurence really serious? I don't want to be the one to break it to them, but theorizing about the mind has been dominated for the last 30 odd years by Putnam's (1975) bizarre Twin Earth example (with matching twins). Moreover, the resultant intuitions from this example led to the widespread belief that the content of psychological states is mainly wide.

So what are Margolis/Laurence trying to say in the above? That when the intuitions from bizarre philosophical inventions give support for wide content they should be accepted, but when the intuitions give support for narrow content they should be disregarded? In my opinion, any intuitions that come from bizarre philosophical inventions should be fairly examined and judged, whether they support narrow content or wide content.

<sup>2</sup> Given the choice, I'd rather be walking down the streets of Boston with Stalnaker, but such is my lot in life.

<sup>3</sup> *The Matrix* is a 1999 science fiction feature film, which stars Keanu Reeves and Laurence Fishburne, and is directed by The Wachowski Brothers.

<sup>4</sup> Loar, 1988b, p.202, gave me the idea for this choice. By "constituted" I mean something like local supervenience, so that a person's narrow notional world is locally supervenient on their neurophysiological states. So if two subjects are neurophysiologically identical then their narrow notional worlds are necessarily identical.

<sup>5</sup> McGinn (1989) claims that "[t]he internal facts that are held constant in such thought experiments divide into three basic sorts: internal states of the body and brain; behavioural dispositions; and proximate stimulations at the sensory receptors" (1989, p.2).

<sup>6</sup> Farkas (2003) also examines the "internal" and "external" distinction between states/properties, etc. that is typically utilised in Twin Earth thought experiments, coming to the conclusion that an identity of internal physical states/properties of subjects is not necessary or sufficient for externalist arguments to go through. This is not really the point that I am making in the text, but my point is related to the notion that Farkas favours for being the relationship shared by subjects in Twin Earth scenarios, namely, that of the "subjective indistinguishability of their situations" (2003, p.196), i.e., the way things seem to the subjects from their first-person viewpoints is identical. This is pretty much the upshot of what I am arguing for in relation to me and my twin, that is, if our internal proximal stimulations are identical, then how things subjectively seem to us should also be identical.

<sup>7</sup> I take something like this to be the reason for Loar to claim that his psychological content only has the appearance or illusion of representationality from the subjective viewpoint

<sup>8</sup> When I have spoken of me and my doppelganger's subjective experiences being objectively simulated, what it should mean is that from our subjective perspectives, that simulation and all the notional objects and states of affairs contained in it, are, so to speak, our respective external worlds. That is, we are not meant to be having experiences of our own experiences, but my doppelganger and I's experiences have been used to construct our notional worlds. The objective simulations of our experiences are the notional worlds or environments that we would expect given those experiences. Perhaps it was something along these lines that Dennett had in mind as well. Recall that Dennett talked of a notional world being a model of one's internal representations, he also goes on to claim that a notional world "...does not consist itself of representations but of representeds. It is the world "I live in", not the world of representations *in me*" (1982, p.40).

<sup>9</sup> My supervisor raised a point here that the aliens might not know this wide content, particularly if they did not have the concepts of gnomes, shops, etc. This is a tricky point, and perhaps all I can say is that I am assuming that the aliens would be able to *perceptually* verify that the wide content and narrow content of my formulated notional world adequately match up. For example, if the aliens froze me while I was on the edge of the Grand Canyon, surveying the view in front me, but when they simulated my narrow notional world it resulted in the Argos shop window and gaudy water-feature being shown, even though they may not have the requisite concepts, they would still be able to tell that there was a mismatch between the narrow and wide contents.

<sup>10</sup> I am not claiming here that narrow content is dependent on wide content. I am just trying to show that, contra Stalnaker, the subject's narrow notional world is not too unconstrained to be workable.

<sup>11</sup> This point that Dennett makes about how internal representations acquire their semantic properties from their relations to things in the environment, is similar to the point that Putnam (1981) makes, where he argues that representations, whether physical or mental, do not have an intrinsic, "magical" connection with what they represent, "a connection independent of how [they are] caused and what the dispositions of the speaker or thinker are" (1981, p.5). Another way to put this is that there are not built-in meanings to words, thoughts, etc., independent of their extrinsic connections to the objects or states of affairs that they are referring to or representing.

<sup>12</sup> This would be analogous to the appearance of intentionality that Loar's psychological content seemed to possess, from the subjective/projective viewpoint of the subject. In saying this, it is odd that Loar (1988b) himself considers that one of the problems of Dennett's approach is that "...narrow content is essentially (not just apparently) intentional..." (1988b, p.202). I am at a loss to explain how Loar comes to this conclusion, as Dennett talks only of a "fictional" or "imaginary" environment.

Moreover, in discussing what beliefs in a notional world would be “about”, Dennett (1982) believes that he can use what is called “Pickwick-aboutness”, which would give some semantic values to the beliefs, but which would not be “genuinely relational” (1982, p.45).

<sup>13</sup> The internal meaning/content that I am trying to describe here could well be what Perlis (1991) describes as “Iref” or internal reference. What Perlis means by this is that, for example, a subject could intend a mental symbol “dog” to refer to the internal representation of a dog, not the external dog that caused the internal representation (1991, p.438). It seems to be only this internal referring that an organismically formulated notional world provides, i.e., when I am referring to the gnome and the water coming from its bottom in my notional world, it is the internal representation of the gnome with water coming from its bottom that I am referring to. However, despite this, I still *mean* or *intend* something, when I internally refer like this.

<sup>14</sup> It is only after I wrote this that I realised that the “representational superimposition” that I was talking about was basically the same as the notion of a virtual reality. This only dawned on me after reading Velmans (1998), where he is trying to find a model of perception that would accommodate a virtual reality. Velmans declares that “To accommodate virtual reality one needs a model of perception that more accurately portrays the *phenomenology* of what is experienced...” (1998, p.49), and to this end he suggests that a “reflexive model” of perception may achieve this, where “...the subject’s experience of a cat is just the *cat as-perceived out in space*. That is, an entity in space, once it engages the visual system, is experienced as an entity in space. That is why the entire process is ‘reflexive’” (1998, p.49). This is pretty much what I am trying to suggest in the notional Argos-episode, my experience of the whole situation seems to have the spatial properties that it would have if it were a veridical experience.

<sup>15</sup> Could Burge’s (1986) argument for an externalist position in relation to representational or perceptual content be used here to attack my claim that me and my twin’s perceptual contents are identical? Briefly, the argument involves considering a situation where a subject’s particular visual perceptual states are normally caused by thin shadows. Then, a counterfactual situation is considered where the subject’s intrinsic physical states (including proximal stimulations, retinal states, etc.) are kept the same, but now the particular visual perceptual states are normally caused by thin cracks (there are no thin shadows in the counterfactual environment). The conclusion that Burge comes to is that despite the identity of intrinsic physical states, the ascriptions of perceptual content to the subject in the actual and counterfactual situations would be different, i.e., in the actual situation the ascribed perceptual content would be “a thin shadow” and in the counterfactual situation the ascribed perceptual content would be “a thin crack”.

Therefore, different distal causal sources seem to result in different perceptual contents, despite identical proximal stimulations on the sensory receptors. Could this argument also apply to the situation of me and my twin? That is, could it be argued that me and my twin’s notional “water” representations do not have identical content, due to the fact that my “water” representations are normally caused by H<sub>2</sub>O and my twin’s “water” representations are normally caused by XYZ, i.e., the content of my “water” is water or H<sub>2</sub>O whilst the content of my twin’s “water” is twater or XYZ. This is a possibility, however, it seems just as plausible to me to posit a common perceptual content that is shared between me and my twin’s water representations; particularly if perceptual or representational content is understood in terms of how things seem to the subject from his first-person perspective. This is the strategy that has been suggested by Matthews (1988), McGinn (1989) and Segal (1989b, 1991) in order to counter Burge’s conclusion. The suggestion is that instead of the different perceptual contents of “a thin shadow” and “a thin crack” there could be a shared perceptual content of “a thin dark line” or something like that (Burge’s position is defended in Davies (1991a) and Edwards (1994)). In a similar fashion, the shared narrow perceptual content of me and my twin’s water representations could be something like “the transparent, odourless, colourless, liquid etc.”

<sup>16</sup> This is the situation that Burge, 1982, p.108 considers, as he wonders what Putnam means when he talks of narrow psychological states. Burge acknowledges that the situation was originally considered by Fodor (1980).

<sup>17</sup> This claim is similar to Jackson’s (1998) “Locke-inspired” description theory of reference, where “...using the pre-analytic or folk term ‘about’: terms like... ‘water’ ...are used by speakers to talk *about* whatever has the properties they associate with the term in question; or, as philosophers of language might say it, a name *T* used by *S* refers to whatever has the properties that *S* associates with *T*” (1998, p.203).

<sup>18</sup> This would mean that the concepts of me and my twin’s “water” would not be natural kind concepts, but perhaps they could be what Fodor (1994) calls visible property concepts, viz., water-appearance concepts (this point is mentioned in Cheng (2002)). Or in a similar vein, we could follow Segal (2000)

and say that me and my twin's "water" terms expressed a "motley" concept, which did not include just a single natural kind, but may have several natural kinds in its extension, i.e., H<sub>2</sub>O and XYZ (2000, p.132).

<sup>19</sup> It seems intuitive to me that this would have been the situation with regards to the meaning/reference of "water" on Earth and Twin Earth in, say, the year 1750, before the different chemical compositions of the respective waters were discovered.

<sup>20</sup> Burge does acknowledge that the examples he uses could be said to involve "relational propositional attitudes" (i.e., *de re* propositional attitudes), and so "...do not strictly show that Adam and Adam(te) differ in their *de dicto* attitudes – attitudes in the narrow sense" (1982, p.111).

<sup>21</sup> This situation seems analogous to Burge's arthritis example, where the obliquely occurring expression "arthritis" in Bert's belief that he has arthritis in his thigh, is claimed to have the socially accepted meaning, i.e., an inflammation of the joints. However, it seems to me that this cannot be providing and characterising the content of Bert's arthritis-belief, at least from Bert's subjective viewpoint, as it seems intuitive to claim that Bert would not assent to having an inflammation of the joints in his thigh.

<sup>22</sup> The points made in this paragraph roughly accord with the view put forward by Stroll (1998), where he is discussing Kripke's and Putnam's essentialist position on the individuation conditions of natural kinds, such as water, and the upshot this has for the semantic values of terms which are used to refer to natural kinds. In Stroll's opinion, contra Putnam, the formula Water = H<sub>2</sub>O is not an identity statement and should not be taken to entail that the term "water" means H<sub>2</sub>O. Stroll takes the view that it is the observable or phenomenal properties of water that in everyday life plays the main role in individuating water and giving the meaning to the term "water" (1998, pp.46-52). This idea that the natural kind itself may not provide the most plausible semantic content for a so-called natural kind term is echoed in Gray (2006), where he discusses the reference of natural phenomenon terms like "light". Kripke (1980) tried to argue that a term like "light" would have its reference fixed by the natural phenomenon it refers to, i.e., "a stream of photons", rather than the original reference which is fixed by a definite description which involves the visual sensations of light (2006, pp.141-143). However, Gray argues persuasively that it is more plausible that the reference of "light" stays fixed by its original definite description and that, effectively, "Light = a stream of photons" is not a genuine identity statement (2006, pp.144-145). This comes from the fact that scientists standardly use the term "light" only to refer to the visible part of the electromagnetic spectrum, whereas, if light was a stream of photons then "light" would refer not only to the visible part, but also to the non-visible parts of the electromagnetic spectrum, e.g., to gamma rays, x-rays, etc., a result which seems implausible (2006, pp.144-146). A rough analogue of this situation in relation to a natural kind term such as "water", would (I think) be that Kripke would be committed to arguing that, if on a possible world, H<sub>2</sub>O is a black tarry substance which is used to make pottery, then our term "water" should correctly refer to this substance. This would be the case even if on this possible world there is a different substance which has all the visible and conceptual properties that water has in the actual world. Of course, this is only a rough analogue, because natural kinds like water have a definite chemical composition, a definite natural division, whereas a natural phenomenon like light does not, but it still seems to me a counterintuitive result. That is, it seems far more intuitive that the rigidified definite description of water from the actual world continues to fix the reference of the term "water" on the counterfactual possible world, rather than the natural kind itself.

<sup>23</sup> It seems to me that Lepore/Loewer are also question-begging against the idea of a narrow content when they argue that "...truly phenomenological expressions" like "is red" are also nonsolipsistic (1986, pp.606-607). The gist of the argument is that things which are/appear red on Earth are actually green on Twin Earth, even though they still appear red to Twin Earthlings. So that if an Earthling came to Twin Earth, the things that he/she would correctly call green (e.g., green roses, green boiled lobsters, etc) are being called red by a Twin Earthling. The upshot of this is that Lepore/Loewer take the view that the Earthling and Twin Earthling "...are thinking different thoughts when each says to herself "That's a red one", even though they are each in neurophysiologically identical states and they are experiencing the same qualia" (1986, p.607). But why be so uncharitable to the idea of a shared narrow content between the twins? It could still be claimed that the twins' thoughts had a different wide content, while allowing that there is an obvious sense in which the thoughts have the same narrow content, i.e., how things seem to the twins is identical. Some sort of shared content would explain why, if the twins were each back on their home planets (i.e., in separate contexts) and were asked to pick out things that they thought were red, they would pick out the same types of things, e.g., roses, boiled lobsters, etc.

## **PART 4**

# **A PROPOSAL FOR AN ADEQUATE AND NECESSARY NARROW CONTENT**

In this part of the thesis I put forward my own preferred account of narrow content, which will hopefully be seen as an adequate and necessary account. However, before attempting this, it will perhaps be useful to take stock of the position that we have reached so far. In the previous three parts we have examined specific examples of what I take to be three of the most popular construals of narrow content, viz., conceptual role, indexical and representational (which consists of perceptual content). Loar's psychological content was the exemplar of conceptual role narrow content, Fodor's mapping theory was the exemplar of indexical narrow content and Dennett's notional worlds was the exemplar of representational narrow content.

Three conditions of adequacy were specified that each construal would have to satisfy in order to be considered an adequate theory of narrow content. What we have found



is that none of the exemplars of the three construals of narrow content have satisfied all three of the adequacy conditions. In this part I will attempt to formulate a theory of narrow content which does satisfy all three conditions of adequacy for narrow content. In chapter 8 I will explicate the theory and deal with criticisms that could be made of it. It will be seen that the theory bears most resemblance to Dennett's notional worlds conception, and so is a representational (or perceptual) narrow content, with mainly visual perception being involved. In chapter 9 I will try to argue that the narrow representational content produced from my theory also provides necessary and sufficient conditions for a subject to entertain a belief with a specific content.

## CHAPTER 8

### THE FORMULATION OF AN ADEQUATE NARROW CONTENT

#### 8.1 Narrow Content as a Subjective Virtual Reality

To begin with, consider the following situation (which is based on a similar situation given in Putnam (1981)):

(i) In the actual world, there is a subject S, who is visually perceiving a glass of water on a table. Because of this, S comes to have the belief that there is a glass of water on the table. In a counterfactual world, there is a subject S\*, who is an internal physical duplicate or twin of S, who is also visually perceiving a glass of water on a table.

Because of this, S\* also comes to have the belief that there is a glass of water on the table. As is usual in these situations, the actual world and counterfactual world are physically identical, except for the fact that water in the actual world is composed of H<sub>2</sub>O and water in the counterfactual world is composed of XYZ.

The above example has both S and S\* having a perceptual experience which then leads to a certain belief based on that perceptual experience.<sup>1</sup> The reason for this is that the narrow content I am going to argue for in this chapter is mainly a representational one, which consists of perceptual content; something along the lines of the narrow content that we had from considering the representational version of

Dennett's notional worlds. It is a narrow content that results from how the subject represents the external world, or equivalently, from how the world seems to the subject.

Because of this, there are going to be a number of presuppositions involved with my argument. Firstly, I am going to presuppose a version of a representational theory of perception, where perceptual experience of the external world is not direct, but consists of neurophysiologically produced representations of the external world. In saying this, I don't mean to imply that in cases of veridical visual perception the subject is consciously aware of these internal representations (i.e., there will still be "transparency" of visual experience), but nevertheless, the subject's perceptual experience will still be understood to involve internal representations.<sup>2</sup> Moreover, the internal representations will be intentional in character, in that they will involve objects and events that are considered to be external to the subject, but they will be internal in that these representations are considered to be produced from the proximal stimulations in, so to speak.

Secondly, I am presupposing a version of a representational theory of the mind, which involves the manipulation and application of mental symbols or tokens to the above posited internal representations. It would then be the concepts that the subject associates with the application of certain mental symbols to certain internal representations that would go towards forming the content of the subject's beliefs and other propositional attitudes.

Thirdly, I am going to presuppose the supervenience of the mental on the physical, or more specifically, the supervenience of the internal representational properties that mental states have on the local neurophysical properties. The upshot of this is that if two subjects are neurophysically identical, then how they phenomenally or qualitatively represent the world to be, will also be identical.<sup>3</sup> Lastly, as before, I will be presupposing that the internal neurophysical properties of the subject include the proximal stimulations on the sensory receptors of the subject.

With the presuppositions out of the way, let us now examine situation (i). The typical externalist view here will be to judge the belief-contents of S and S\* as different, due to the different chemical compositions of the respective “water” natural kinds involved. S believes that there is a glass of H<sub>2</sub>O/water on the table, while S\* believes that there is a glass of XYZ/twater on the table. Moreover, this leads to the view that the beliefs qua beliefs are also different, due to the fact that they have different truth conditions. S’s belief is true if and only if a glass of H<sub>2</sub>O is on the table and S\*’s belief is true if and only if a glass of XYZ is on the table. However, as we have seen, a narrow content theorist could accept these externalist conclusions but still posit a narrow content that is shared between S and S\*. It could be claimed that while the wide truth conditional content of the beliefs are different, there could still be a narrow content that the beliefs shared.<sup>4</sup> It is to the formulation of that narrow content that we now turn.

In order to formulate the narrow content that is shared between S and S\* we need to ascertain the contribution that their internal physical properties make to their respective belief-contents; or to put it in Dennett’s terms, we need to ascertain what

the organismic contribution is to their respective propositional attitudes. Therefore, as before, imagine that S and S\* are frozen mid-experience and mid-believing that there is a glass of water on the table. They are then attached to the alien supercomputer of the previous part, which, only utilising their internal physical properties, produces a 3-D projection of what S and S\* are experiencing from their subjective viewpoints.<sup>5</sup>

The internal physical properties include the proximal stimulations on the visual sensory receptors, the resulting electrical signals to the brain, where cognitive or perceptual processing then produces the neural representations which are subjectively experienced by the subjects.<sup>6</sup> The intuitive upshot of all this, is that S and S\* will be having identical representational experiences of perceiving a glass of water on a table. That is, how they perceptually “picture” things to be will be the same. Of course, in the standard Twin Earth thought experiments, the isomorphism of how things phenomenally seem to the twins is usually a given, but I just wanted to make sure that it is understood this given is based on the isomorphism of the twins’ internal physical properties. This amounts to the claim that S and S\*’s internal representational states are locally supervenient on their brain states.

I previously claimed that what I had in mind was that the internal representational experiences of a person were composed of the surface appearances of the external world, like a representational superimposition of the external world, but without the physical external world actually being under the superimposition. The notion that I was trying to describe I later realised was that of a *subjective virtual reality*, which has the appearance of being located and extended in space.<sup>7</sup> Therefore, in situation (i), due to their identical internal neurophysical states, S and S\* share a narrow

representational content that results in them having identical subjective virtual realities of that situation.

I suppose that what I am suggesting here has a phenomenistic feel to it, where the subject only immediately apprehends the internal representations that result from the proximal stimulations on his visual system. This is a fair point, as the view of perceptual relations to the external world that seems to be emerging here is that of external objects and the light they reflect being causally responsible for the proximal stimulations on the visual receptors of the subject, which then produce the internal representations by which the subject can “picture” the external objects. Of course, what we have with the shared narrow representational content is just the internal part of this relation, without the external reality.

### 3.2 The Semantic Properties of Narrow Representational Content

However, it is important to note that what we have so far is a narrow *representational* content shared between S and S\*; but does this translate into a narrow *semantic* content that is shared between S and S\*? What I mean by this can be illustrated by Jackendoff’s (1991) claim that there are two versions of “...the fundamental question for a theory of mind” (Jackendoff, 1991, p.411). There is a “psychological” version of the question which deals with issues like perception and how the brain constructs reality so that the world seems the way it does to us (Jackendoff, 1991, p.412). The other version of the question is a “philosophical” one, which deals with issues like meaning, truth and reference (Jackendoff, 1991, p.411). So far in this chapter we

have only dealt with the “psychological” issues of the proposed narrow content.

Therefore, it is now time to examine the “philosophical” issues of meaning, truth and reference in relation to the narrow representational content that we have so far.

### 8.2.1 *Narrow Meaning/Content Properties*

To recap, the situation that we have is that S and S\* are attached to the alien supercomputer, which, utilising only their internal physical properties (from the proximal stimulations in, so to speak), is objectively projecting what S and S\* are subjectively experiencing. Intuitively, the result of this is that S and S\* have identical internal representations of their situations, which means that they are subjectively apprehending the same represented scene, i.e., that of a table with a glass of water on it. Therefore, the question now is what narrow semantic content or meaning is shared between S and S\*’s respective beliefs that there is a glass of water on the table; or to put it another way, are S and S\*’s belief-state tokens of the same semantic belief type.

To begin with, let us return to situation (i) which is the veridical situation for both S and S\*, i.e., they are both perceiving the distal or external world glass of water on the table. One could describe this as that S/S\* are denoting the distal objects of the situation using certain mental symbols. The distal table is denoted with the symbol “Table”, the distal glass is denoted with the symbol “Glass” and the distal water is denoted with the symbol “Water”. Of course, in describing this denotation of S/S\*, I don’t mean that they have a conscious awareness of denoting, deliberately aiming their mental symbols at the distal objects. It is more plausible to imagine the process

of denotation in situation (i) as occurring instantaneously, almost at an unconscious level. We can imagine that both S and S\* immediately recognise the distal objects in front of them and that the process of denotation is also immediate. It is then from this almost instantaneous recognition and denotation in situation (i), that S and S\* come to acquire their respective beliefs that there is a glass of water on the table.

Now, what is the intuition regarding what S and S\* are experiencing whilst connected to the alien supercomputer? If what I have previously claimed is accepted, then the intuition will be that S and S\* are subjectively experiencing a qualitatively identical virtual reality (nonveridical) version of the veridical situation (i). How things seem to S and S\* whilst connected to the alien supercomputer is that they are perceiving a glass of water on the table, which has the appearance of being spatially located and extended. Moreover, I would further claim that it is plausible to imagine that from their shared narrow virtual realities, both S and S\* could still come to acquire the belief that there is a glass of water on the table. That is, there could still be instantaneous recognition and denotation that occurred, which then led to the formation of the belief.

Of course, whilst attached to the alien supercomputer, S/S\* would not be denoting the distal table, glass and water, but rather the internal representations of those distal objects, which have been cognitively processed from the proximal stimulations in, i.e., by only using the internal physical states and processes of S and S\*. <sup>8</sup> Moreover, it is important to note here that S and S\* would not know that they are denoting the internal virtual representations of the distal objects from situation (i); from S and S\*'s subjective viewpoints they would still be appearing to denote external world objects.



Therefore, what I am claiming with the above can now be put as follows. The narrow semantic content that S and S\* share from their qualitatively identical narrow representational content is almost the same as the semantic content that they share whilst in the veridical situation (i). The narrow semantic content of their shared virtual realities can go towards providing S and S\* with enough meaning to enable them to form their shared consequent beliefs that there is a glass of water on the table. This comes from the fact that S and S\* both denote the same respective internal representations with the same symbols, i.e., the mental symbol “Table” refers to the internal representation [Table], the mental symbol “Glass” refers to the internal representation [Glass] and the mental symbol “Water” refers to the internal representation [Water].<sup>9</sup>

**(An important point of clarification should be made here. When I talk of internal representations like [Table], [Glass], [Water], etc., I do not mean that the representations have the syntactic shape of the words used in the brackets. The internal representations are composed of the virtual images of the table, glass, water, etc. I just use the words in the brackets as a method of specifying what virtual image is being referred to by the subject. For example, [Table] stands for the virtual image of a table, which will be what ever sort of table is involved in situation (i), perhaps it will be an image of a wooden table with four legs or it could be an image of a metal table with one central supporting leg. The point is that the internal representations are composed from the visual phenomenology of the subject, how things visually seem to the subject.)**

Of course, when it is claimed that the semantic content of the nonveridical situation (i) is “almost the same” as the semantic content of the veridical situation (i), what this means is the following. The semantic content from the veridical situation (i) enables one to go to a greater semantic depth into the external world than the semantic content from the nonveridical situation (i). By this I mean that in the veridical situation (i), one can differentiate between the actual H<sub>2</sub>O and the counterfactual XYZ which constitutes the respective waters; the wide semantic content enables one to make a distinction in reality at the microstructural level. This is not possible in the nonveridical situation (i), where the narrow semantic content only goes to a certain depth into the world; roughly, to the depth where the actual and counterfactual waters are both specified by the more general description, say, transparent liquid. Still, this seems to me to be quite a reasonable result, considering that the narrow semantic content in question has been produced only using the internal physical properties of S and S\*.

With regards to the question of what narrow semantic content the postulated narrow representational content provides, I believe that the above is a satisfactory account. Moreover, I think the above also gives an answer to the question of what constitutes the mental representations themselves. The internal representations which are produced by the internal neurophysical properties of the subject take the form of a “visual” experience of the situation involved, which from the first-person perspective of the subject seems to be spatially located and extended, even though it is only a virtual or notional experience of the situation.<sup>10</sup>

With regards to the question of what makes the internal representation of situation (i) *about* situation (i), I don't think I need answer this question. From my description of the internal virtual "visual" experiences of S and S\*, the representational relationship which connects those virtual experiences to a certain distal state of affairs, is perhaps a causal theory of representation. However, I feel that as far as the issue of narrow representational content goes, it is not necessary for me to also specify and argue for a particular representational relationship that that narrow representational content has to the external world.<sup>11</sup> What I was trying to work out is what semantic content comes from that narrow representational content, and this is what I believe I have done with the above. The narrow semantic content provided by S and S\*'s narrow (i.e., nonveridical) representational content of situation (i) is similar to the semantic content that is provided by S and S\*'s wide (i.e., veridical) representational content of situation (i).

Moreover, with regards to specifying the narrow content shared between S and S\*, it seems plausible that one can just specify the particular concepts related to the particular mental symbols. With the symbols "Table" and "Glass" this should not be problematic, as there is no difference between these concepts in the actual and counterfactual situations. However, is there a problem with specifying the narrow content of the respective "Water" symbols that S and S\* use? After all, it could be claimed that the concepts associated with the "Water" symbols are different in the actual and counterfactual situations, i.e., the concept of actual "Water" is H<sub>2</sub>O, whilst the concept of counterfactual "Water" is XYZ.

It seems to me that this difference does not cause a problem to specifying the narrow meaning of S and S\*'s "Water" symbols. Certainly, one could plausibly claim that the wide content of the respective "Water" symbols is distinct, due to the distinctness of the chemical compositions of the substances involved. However, in gauging the narrow content, we are only using what is given in the internal representations (and also narrow conceptual roles) involved with the respective "Water" symbol references of S and S\*. Therefore, it seems plausible to say that the shared concept of the respective "Water" symbols is the stereotype or definite description of water. Perhaps one could give a specification of this with the following:  $x = \text{water}$ , if  $x$  is a transparent, colourless, odourless, etc. liquid in one's local environment. Therefore, it is perhaps more accurate to describe the reference of the symbol "Water" as referring to, not [Water], but to something like [the transparent, colourless, odourless, etc., liquid]. It is this latter specification of the internal representation that is shared between S and S\*, that best describes the shared narrow meaning that is attached to their respective "Water" symbols.

The upshot of this is that with situation (i), there is a plausible narrow meaning that is shared between S and S\*'s respective beliefs that there is a glass of water on the table. Moreover, this narrow meaning could be specified in the sentence "S/S\* believes(sn) that there is a glass of water on the table", where the "sn" signifies a subjective (notional) de dicto belief ascription.<sup>12</sup> It is also worth remembering that all this has been achieved only using the internal physical properties of S and S\*. The key to this result is the inclusion of the proximal stimulations on the subject's sensory receptors in the inventory of internal physical properties. So even though the resulting subjective virtual reality contains things which are supposedly outside the body of the

subject, these things are still considered to be internal as they are constituted only by the neurophysical events and processes in the subject's body.

The situation could be described like this: even though from the objective third-person point of view the brain is only a collection of neurophysical states and processes (perhaps a "syntactic engine"?), from the subjective first-person point of view of the brain's owner, the result of all this neurophysical activity is a virtual world built from internal representations and which provides a "semantic engine" for the subject's use. Moreover, it can also be claimed that because it is from the subjective first-person viewpoint that a person apprehends his or her own consciousness, this consciousness would automatically accompany the internal referring, the upshot of which would be that the person is having beliefs, desires, etc. that are contained in his or her own conscious experience.<sup>13</sup>

### 8.2.2 *Narrow Truth Conditional Properties*

Having examined what meaning the proposed narrow representational content could provide, it is now time to consider what truth conditional properties this narrow content contains. Answering this question will basically involve examining how, given the narrow representational content, this content enables the subject to have intentional or world-directed thoughts and beliefs. The truth conditions of a belief are usually defined as being a certain situation in the world that has to obtain in order for the particular belief to be true. For example, the truth conditions of the belief that

snow is white will be that snow is white. If snow is white then the belief is true, if snow is not white then the belief is false.

From the narrow representational content that is shared by S and S\*, it has been proposed that S and S\* also share the narrow subjective de dicto belief that there is a glass of water on the table. From what has just been said, it seems straightforward that the truth conditions of this shared narrow belief will be that there is a glass of water on the table. However, one must be careful at this point, to make quite clear that it is the *narrow truth conditional content* of the shared belief that we are interested in. What is usually called the *wide truth conditional content* of the shared belief, will, on the usual understanding, result in the shared belief having different truth conditions and so not being a shared belief at all. That is, on a wide reading, the truth conditions of S's belief will be that there is a glass of H<sub>2</sub>O on the table, whilst the truth conditions of S\*'s belief will be that there is a glass of XYZ on the table.

Therefore, it is the narrow truth conditional content of S and S\*'s shared belief that we are interested in. It seems plausible that the narrow truth conditions of the shared belief that there is a glass of water on the table will have to come from how things seem to S and S\* from their subjective viewpoints, i.e., from their identical internal representations of the external world. To help achieve this, we could use Loar's notion of "context-indeterminate realization conditions" or Jackson's (2003) notion of "God models". With Loar's notion, the context-indeterminate realization conditions of a belief are the set of possible worlds in which that belief would be true (and not a misconception). With the narrow belief of S and S\*, this would result in a possible world where there is a glass of transparent liquid called water on a table, which, if it is

realized in the actual world, would be composed of H<sub>2</sub>O, or if it is realized in the counterfactual world, would be composed of XYZ. With Jackson's notion, it is imagined that God has the power to model the world external to the subject, with the result that the finished model would make the subject's belief true, but true from the subject's point of view (Jackson, 2003, p.101). Therefore, with S and S\*'s narrow belief that there is a glass of water on the table, to make this belief true from the subjective perspective, all God would have to do is build a model of the world where there is a transparent liquid called water in a glass which is on a table. God would not need to concern itself with the microstructure of the water, as this fact is not salient to the truth of the belief from the subjective perspectives of S and S\*. So, as with Loar's realization conditions, the microstructure of the water could either be H<sub>2</sub>O or XYZ, it would not matter to the narrow truth conditions of the belief.

Therefore, having described the notions of Loar and Jackson, are they what is required to specify the narrow truth conditional content of S and S\*'s shared belief that there is a glass of water on the table? I think that they are, as it was suggested above that the narrow truth conditional content of S and S\*'s belief should depend only on the identical internal representational properties that are available from their respective subjective viewpoints, and it this notion that seems integral to the notions of Loar and Jackson. That is, with both notions, the individuation conditions of the "Water" concerned is that it is a transparent, colourless, odourless, etc. liquid, which is how situation (i) is internally represented from the subjective perspectives of S and S\*.

Therefore, in situation (i), we can say that the (narrow) truth conditions of the narrow belief that there is a glass of water on the table are that it is true if and only if the transparent liquid contained in the glass on the table is water. As with the notions of Loar and Jackson, it doesn't matter to the narrow truth conditions whether the microstructures of the actual and counterfactual waters are different, the important factor is that in the actual and counterfactual scenarios of situation (i), the transparent, colourless, odourless, etc. liquid in the glass on the table is called water in that environment.

#### An Important Objection to the Above Construal of Narrow Truth Conditions

Before moving on, I would like to clarify something about the above. An objection could be made here that the narrow truth conditions of S and S\*'s beliefs would be too "open", due to them being individuated solely by representational or perceptual properties. Transparent, colourless, liquids like 7-UP, acid, etc. could also produce the same narrow representational states in S and S\* as the respective waters produced in situation (i). Therefore, how would one be able to narrowly discriminate between water, 7-UP and acid?

A plausible answer to this question would be to bring in the narrow conceptual roles associated with the internal [transparent liquid] representation that is associated with the "Water" symbol, and use these to individuate that internal representation more finely. For example, the narrow concept associated with [transparent liquid] could include the following: falls from the sky and fills the oceans and rivers, comes out of



taps, is good for sailing on, etc. Conceptual roles such as these would plausibly enable one to discriminate between which [transparent liquid] was being internally referred to by “Water”. Moreover, because S and S\* are considered to be functional duplicates, they would have identical narrow conceptual roles for [transparent liquid]; which means that they would have identical discriminatory powers for distinguishing various transparent liquids. Therefore, by utilising the narrow “Water” conceptual roles associated with the internal representation [transparent liquid], it seems plausible that S and S\* would be able to individuate it finely enough to discount a situation where the narrow truth conditions of their belief that there is a glass of water on the table also included other transparent liquids such as 7-UP, acid, etc.

The above suggestion sounds reasonable enough, but there is a more straightforward way of avoiding the above objection. The important point is that the proposed narrow truth conditions are supposed to apply to the narrow *belief* which is shared between S and S\*, i.e., the narrow belief that there is a glass of *water* on the table. I have italicised the words “belief” and “water” to emphasise the point that it is a belief concerning water that the narrow truth conditions apply to. Of course, we are supposing that it is only from the internal, narrow, representational states of S and S\* that this narrow belief has been produced, so the upshot of this will be that there is an internal representation of a glass containing a transparent liquid on the table. Moreover, this transparent liquid that is internally represented could be a number of different liquids such as water/twater, 7-UP or acid, that is, one cannot discriminate between them just given the internal representation of [transparent liquid].

However, to reiterate the above point, it is not the narrow representational content itself that the narrow truth conditions apply to, it is the consequent narrow belief that comes from the narrow representational content. If this is accepted, then the narrow belief of S and S\* concerns a transparent liquid which is called water, which has resulted from S/S\* both referring to the internal representation [transparent liquid] with a “Water” symbol. (Having said the above, it seems plausible to me to claim that the way S/S\* are *narrowly representing* things in situation (i), i.e., how things seem to them from their subjective viewpoints, are that there is a glass of transparent liquid called water on the table. The point is that S/S\* are not narrowly representing a state of affairs where there is a glass of transparent liquid on the table, they have given that transparent liquid a name, by denoting the [transparent liquid] representation with the “Water” symbol).

The upshot of this is that the narrow truth conditions for S/S\*’s narrow belief that there is a glass of water on the table will depend on there being a transparent liquid called water in the glass on the table. Therefore, it is the narrow belief itself that seems to provide the discriminatory powers that are fine-grained enough to distinguish between various qualitatively identical transparent liquids that may be in the glass on the table. For example, in situation (i), if it turned out that the actual and counterfactual transparent liquid in the glass is 7-UP, then the shared narrow belief of S and S\* will be false. Even though a glass of 7-UP and a glass of water produce a qualitatively identical internal representation [transparent liquid], because the consequent narrow belief of S and S\* specifically concerns a transparent liquid called water, this puts a constraint on the narrow truth conditions of that narrow belief

Perhaps an objection to the above idea could be made here, which is similar in nature to the earlier objection of Block (1991) against the narrow determination of the intension or character involved with Fodor's mapping theory. It could be claimed that in the counterfactual scenario of situation (i) the context of acquisition of the word "water" could be different to the context of acquisition for "water" in the actual scenario. For example, in the counterfactual scenario, "water" could be conventionally used to refer to, say, geraniums, so that for S\*, his "Water" symbol would be representing something completely different to what it does for S in the actual scenario.<sup>14</sup> The upshot of this would be that S\*'s consequent belief that there is a glass of water on the table would involve a narrow truth condition, which, while being syntactically identical to the narrow truth condition of S's belief, would actually involve the occurrence of a geranium in the glass.

However, I don't think that this is a feasible objection against my above suggestion. As I argued similarly against Block's earlier objection, because it is representational narrow content that is involved, that is, how things in the world seem to be to the subject, this sets a constraint on exactly what is being narrowly represented. With S in the actual scenario of situation (i), his "Water" symbol is referring to the internal representation [transparent liquid]. Therefore, as S\* is an internal physical/functional duplicate of S, the latter's internal representation [transparent liquid] "fixes" the representation that S\*'s "Water" symbol is going to internally refer to, i.e., it is also going to be [transparent liquid]. Therefore, S and S\*'s shared narrow belief that there is a glass of water on the table is going to have the same narrow truth conditions, that is, the belief is true if and only if the transparent liquid in the glass is called water in that environment.

Nevertheless, the above suggestion that S and S\*'s narrow belief itself forms a constraint on the antecedent narrow [transparent liquid] representations of S and S\*, would still allow a counterfactual scenario where, say, 7-UP is the transparent liquid called water in that environment. We could imagine that 7-UP is the transparent, colourless, liquid which comes out of taps, falls from the sky, fills the oceans, etc. In a scenario like that S\*'s narrow belief that there is a glass of water on the table would be true if and only if there is a glass of 7-UP on the table. That is, S\*'s "Water" symbol would be fixed onto 7-UP in the external world.

But what would this situation mean exactly? Presumably, it could be claimed that the "Water" symbol shared by S and S\* would have a disjunctive wide content, as both H<sub>2</sub>O and 7-UP would be in the extension of "Water". However, it also seems plausible to claim that as situation (i) is set up, the "Water" symbol shared by S and S\* would also have a disjunctive wide content, with both H<sub>2</sub>O and XYZ being in the extension of "Water". So in both situations it seems that the narrow "Water" symbol will have a disjunctive wide content. The upshot of this seems to be that it has to be admitted that any distal liquid which satisfies the narrow- water belief/representational/conceptual role constraints of the internal [transparent liquid] will be legitimately contained in the extension of the "Water" symbol.<sup>15</sup>

However, the point that I'm crawling towards is that as situation (i) is set up, this does not mean that the narrow- water belief/representational/conceptual properties associated with the internal [transparent liquid] would be unable to discriminate between a glass of water (whether H<sub>2</sub>O or XYZ) and a glass of 7-UP. Therefore, it

still seems plausible to me that by using this method the important objection to narrow truth conditions for S/S\*'s narrow belief could be avoided.

### A Relevant Example From The Literature

Having given the above explication of how the 7-UP objection to narrow truth conditions might be avoided, as this is such an important issue for the construal of narrow representational content being put forward, perhaps it would be useful to see how the explication would apply to a specific example taken from the philosophical literature. The example I have in mind is from Baker (1987a), which she takes to be a general argument against the coherence of any type of narrow content (Baker, 1987a, p.208). Baker reaches this conclusion by first specifying two constraints, one semantic(S) and one physical(P), which she thinks that a psychologically and scientifically plausible/coherent concept of narrow content should satisfy (Baker, 1987a, p.198). The two constraints are:

(S) (Narrow content + context) > Truth conditions of beliefs

and

(P) If C is a narrow content, and S has a belief with narrow content C, and S' is a molecule-for-molecule duplicate of S, then S' has a belief with narrow content C (Baker, 1987a, p.198).

Baker then claims that:

Any concept of narrow content that jointly satisfied both (S) and (P) would have this consequence: if molecular replicas were in the same contexts, their beliefs would have the same truth conditions (Baker, 1987a, p.198).

This point is important because Baker is going to use it to claim that no concept of narrow content can satisfy both (S) and (P). The thought experiment that will be examined is claimed by Baker to be a situation where molecular replicas are in the same context, and so should have beliefs with identical narrow contents according to (P), but their respective beliefs actually have different truth conditions, which means that (S) is not satisfied, and that their narrow contents must be different (Baker, 1987a, pp.206-207).

The thought experiment can be summarised as follows.<sup>16</sup> It is imagined that there are two girls from two different countries attending the same embassy party. The girls are considered to be molecular duplicates and to speak identical English languages, except for one difference. With one girl, in the country she comes from, they use the word “vodka” to denote gin, whereas with the other girl, in the country she comes from, they use the word “vodka” to denote vodka (as we normally do). Because of this difference, Baker calls the latter girl the English speaker and the former girl the non-English speaker. It is then imagined that both girls see “a clear liquid being served in shot glasses” (Baker, 1987a, 207) which results in both of them pointing towards the glasses and uttering the sentence “Vodka is good to drink”.

The upshot of this as far as Baker is concerned, is that the girls are saying different things when they utter their sentences, viz., the English speaker is saying that vodka is good to drink, whilst the non-English speaker is saying that gin is good to drink (Baker, 1987a, p.207). If the utterances are considered to express the girls’ beliefs, then the consequence of this is that the beliefs seem to have different truth conditions,

with the English speaker's belief being true if and only if vodka is good to drink, and the non-English speaker's belief being true if and only if gin is good to drink. As explained above, Baker takes this difference of truth conditions of the replica-girls' beliefs, who are of course in the same context, to show that they have different narrow contents. That is, even though according to (P) the girls' beliefs should have the same narrow content, they are failing to satisfy (S), which means that they don't have beliefs with the same narrow content (Baker, 1987a, p.208).

One can see that with the above there is a certain similarity to the 7-UP objection to the notion of narrow truth conditions for narrow representational content which has just been examined. Therefore, can the explication given in answer to the 7-UP objection also be utilised here in order to escape the conclusions of Baker's argument? Well, it seems plausible to claim that Baker's version of events given above could be seen as providing the *wide* truth conditions of the replica-girls' respective beliefs that vodka is good to drink. The supporter of narrow content could agree with Baker that the wide truth conditions of the girls' beliefs are different, i.e., the truth of the English speaker's belief will depend on vodka being in the observed glasses, whilst the truth of the non-English speaker's belief will depend on gin being in the observed glasses. Moreover, the upshot of this will be that Baker is quite correct in her claim that the semantic constraint (S) will not be satisfied, and so the girls' beliefs will be judged to have different narrow contents.

However, is it really plausible that this is the end of the semantic road for any specification of narrow content that could possibly be given in Baker's thought experiment? It seems to me that there is a possibility for a representational narrow

content to provide some sort of semantic value that is shared between the two girls' beliefs. To help explicate this notion, let us examine a further quote from Baker, where she is claiming that the girls could share the same nonintentional histories:

In particular, the episodes in which each acquires her beliefs may merit the same nonintentional descriptions. Suppose that the English speaker's beliefs about vodka have their origin in a training session for children of diplomats in which she is presented with a picture of a glass filled with a clear liquid and told that it is vodka and good to drink at receptions. Likewise, the non-English speaker's beliefs about gin have their origin in a training session in which she is presented with a picture of a glass filled with a clear liquid *and told that it is gin* [my italics] and good to drink at receptions. The pictorial episodes by which each girl acquires her attitudes may satisfy the same nonintentional descriptions; the pictures of the glass filled with a clear liquid may be indistinguishable. Both teachers may point to the pictures and emit the same sequence of sounds (Baker, 1987a, pp.207-208).

I may be missing something here, but it seems to me that in the above quote, Baker has provided a concise description which could go towards giving the girls' beliefs a shared semantic content, which in turn would go towards giving the girls' beliefs shared narrow truth conditions in the same context. To begin with, Baker talks of nonintentional descriptions, which I take to be equivalent to the notion of the narrow representations we have been talking about in this chapter. That is, in relation to the present thought experiment, the narrow representations are considered from the proximal stimulations in, which means that in the above training session, the girls have shared internal representations of a glass containing a clear liquid called vodka which is good to drink at receptions. As far as I am concerned, these shared narrow representations have a semantic content, they could go out into the world, but only to a certain semantic depth, i.e., the depth where there is a clear liquid called vodka, and not the semantic depth where the clear liquids are differentiated as being vodka and gin.



This narrow reading of the above quote seems the most plausible to me. Both girls have the same internal representations of a clear liquid in a glass, which is called vodka and is good to drink at receptions. This is why I italicised the section of words in the above quote. Given the way that Baker has set up the thought experiment, it seems just plain wrong that the non-English speaker is told that the clear liquid in the glass is gin. The non-English speaker is surely told that the clear liquid in the glass is called vodka, even though it is in fact gin. But this latter point does not alter the fact that from the subjective first-person perspective of the non-English speaking girl, she comes to believe that the clear liquid in the glass is called vodka.

Therefore, let us now return to the thought experiment, to the point where both girls are gesturing towards the glasses containing the clear liquid, and are uttering their respective sentences “Vodka is good to drink”. Let us accept that they are both considered to be in the same context. The question now is what exactly are the contents of the girls’ beliefs and can those belief-contents share some sort of truth conditions. From what has just been said, it seems plausible to me that both girls are internally representing a state of affairs where there are glasses containing a clear liquid which is good to drink at receptions, and is called vodka. That is, they are both referring to the internal representation [clear liquid which is good to drink at receptions] with “Vodka” mental symbols. The upshot of this is that the shared narrow content of their consequent beliefs will be that “Vodka is good to drink”, where “vodka” means “the clear liquid which is good to drink at receptions”. The upshot of this is that the shared narrow truth conditions of the girls’ beliefs will be that they are true if and only if a clear liquid called vodka is good to drink.

The position that I am arguing for does have some important consequences. It does bring up an issue that will be examined in more detail in the next sub-section on the referential properties of the proposed narrow content, but I think the following can be said at this time. The narrow truth conditions of a belief can be intentional or world-directed, in the sense that they go out into the world to a certain semantic depth. However, this semantic depth is not sufficient to denote a specific or singular object or state of affairs. By this I mean that in the above thought experiment, the narrow truth conditions of the girls' beliefs will be satisfied by whatever clear liquid is called vodka and is good to drink at receptions. However, the clear liquid in question does not have to be vodka qua vodka, it could be gin, 7-UP, water, etc., as long as that liquid is called "vodka" and is good to drink at receptions. In contrast, the wide truth conditions of the girls' beliefs will denote a specific or singular liquid, vodka for the English speaking girl and gin for the non-English speaking girl, which will need to be the specific clear liquid in question that is good to drink at receptions in order to satisfy the wide truth conditions. Therefore, perhaps one could describe the situation as the following: the wide truth conditions of a belief involve the relevant singular external world referents that would satisfy the belief, whereas, the narrow truth conditions of a belief involve the relevant non-singular external world referents that would satisfy the belief.

Nevertheless, having said all that, what are the motivations for my positing the same narrow truth conditions for the girls' beliefs that vodka is good to drink? Well, as before in this thesis, it seems to me that by examining the subjective first-person perspective of the person having a belief, one sometimes gets a more accurate picture

of what content the belief has, at least as far as the subject is concerned. This is in contrast to taking an objective third-person perspective of the person having a belief, which usually results in the belief-content ascribed to that person being the content that the ascriber thinks the belief should have, given the physical/social environment of the person.

For example, with Baker's thought experiment, if we examine the first-person perspective of the English speaking girl when she has the belief she expresses with the utterance "Vodka is good to drink", it seems plausible to me that how she narrowly represents things as being is that there is a clear liquid called vodka which is good to drink at receptions. Therefore, the narrow truth conditions of her belief would be that it is true if and only if a clear liquid called vodka is good to drink. In the case of the first-person perspective of the non-English speaking girl, when she has the belief she expresses with the utterance "Vodka is good to drink", it again seems plausible that how she represents things as being is that there is a clear liquid called vodka which is good to drink at receptions. Therefore, once again, the narrow truth conditions of her belief would be that it is true if and only if a clear liquid called vodka is good to drink.

Perhaps at this point it could be claimed that I am begging the question against Baker's position, which is that it is the wide truth conditions of the girls' beliefs which are relevant here, with the result that the truth conditions for the non-English speaking girl's belief that vodka is good to drink are different to the English speaking girl's belief that vodka is good to drink. That is, the former's belief will have truth conditions involving gin whilst the latter's will have truth conditions involving vodka.

However, in support of my position, I would again ask which is the most plausible, that the content of the non-English speaking girl's belief is "a clear liquid called vodka is good to drink at receptions" or "a clear liquid called gin is good to drink at receptions"? This seems to be a "no contest" if we use the disquotational principle of belief ascription with the non-English speaking girl. That is, it seems much more plausible that the non-English speaking girl will sincerely assent to the sentence "You believe that a clear liquid called vodka is good to drink at receptions" and will *not* sincerely assent to the sentence "You believe that a clear liquid called gin is good to drink at receptions". To me, this just seems to be the common sense result that naturally comes from the structure of Baker's thought experiment. The non-English speaking girl just does not know that the clear liquid called vodka is actually gin, so how can gin plausibly be a part of her belief-content? Moreover, how can gin play a role in explaining her behaviour at the reception? It again seems plausible that the non-English speaking girl would explain her behaviour of gesturing towards the glasses of clear liquid and making her utterance by claiming that she believed the clear liquid in the glasses was vodka. It seems totally implausible to suppose that she would explain her behaviour by claiming that she believed the clear liquid in the glasses was gin.

It seems to me that the only way for Baker to escape the above plausibility of the girls' beliefs sharing narrow truth conditions is to play on the fact that the non-English speaking girl's "vodka" term is directly translatable into, and in fact *means*, gin; and this is indeed what Baker tries to do. While describing the thought experiment, Baker claims that:

Assume that translators have established that the word in the remote language that sounds like the English word 'vodka' should be consistently translated as 'gin'...Just as the Spanish word 'burro' means 'butter', so the English word 'vodka' means vodka and the mythical-language word 'vodka' means gin (Baker, 1987a, p.207).

But surely this is question-begging against the supporter of narrow content, Baker is trying to have her wide cake and eat it here. It seems that the reason Baker believes there is only wide and different truth conditions for the girls' beliefs is basically because that is what she has stipulated in setting up the thought experiment. That is, she has stipulated that the English speaking girl's "vodka" means vodka and that the non-English speaking girl's "vodka" means gin and that is the end of the semantic story. If "vodka" means vodka to the English speaking girl and "vodka" means gin to the non-English speaking girl, then there is no possibility of any narrow content that could be shared between the girls' beliefs.

If Baker's thought experiment is to have any non-question-begging plausibility, it must surely be allowed that both girls do not have the semantic ability to distinguish between vodka and gin. If this is allowed, then a semantic gap opens for the possibility of some sort of narrow content along the lines of "the clear liquid called "vodka"", which could be shared between the girls' beliefs that vodka is good to drink. Indeed, Baker brings up this possibility when discussing a suggestion made by Fred Feldman (Baker, 1987a, p.210). Her answer to this is that "Since, however, we typically acquire beliefs in the fashion described and we typically lack omniscience, I see no credible way of denying that the girls have the [wide vodka and gin] beliefs ascribed" (Baker, 1987a, p.211).

However, if Baker could stop begging the question for one moment, she would see that there is no need to deny that the girls' have the beliefs she is claiming. The supporter of narrow content can quite happily accept that the wide truth conditions of the girls' beliefs that vodka is good to drink are different, i.e., that the girls' wide beliefs are different. But this does not mean that there cannot also be a plausible specification of narrow truth conditions for the girls' beliefs, i.e., that the girls' narrow beliefs are the same. Indeed, it is ironic that Baker should say that we lack omniscience, for if, for example, the non-English speaking girl only knows that a clear liquid is called vodka, then ascribing her the belief that the clear liquid is actually gin seems to require omniscience on the part of the non-English speaking girl to actually know this.

Therefore, it seems to me that, contra Baker, there can be a plausible narrow content which satisfies the constraints (S) and (P), as long as it is accepted that the truth conditions in (S) are narrow in nature. On a wide construal, the truth conditions of the girls' beliefs are distinct, which indicates that the contents of their beliefs are type-distinct in the same context. However, on a narrow construal, the truth conditions of the girls' beliefs are the same, which indicates that the contents of their beliefs are type-identical in the same context.<sup>17</sup> Furthermore, this is hopefully seen as another example of how the postulated narrow representational content can provide beliefs with narrow truth conditions which are intentional and have certain discriminatory powers.

### 8.2.3 *Narrow Referential Properties*

Finally, we come to the question of what referential properties the narrow belief that there is a glass of water on the table has. It is worth remembering that the narrow belief is only constructed from the narrow representational content of situation (i), which from the subjective perspectives of S and S\*, formed a virtual reality which presented a situation where there is a glass of water on the table apparently extended and located in space. It was suggested that the subject denotes these internal representations with specific mental symbols, e.g., the subject would, almost instantaneously and unconsciously, use a “Table” symbol to refer to the internal representation of a [Table]. Therefore, it is from this internal representation and reference that we must now attempt to move out to the external world referents of the narrow belief that there is a glass of water on the table.

So far in this explication of narrow content, we have brought in elements from the two previously examined construals of narrow content, namely, narrow representational states and conceptual roles.<sup>18</sup> Therefore, it seems appropriate here that we bring in an element from the third narrow content construal which was the indexical one involving Fodor’s mapping theory. The element I have in mind is the *intension* of the narrow belief content that there is a glass of water on the table, formed from the combination of narrow representational states and conceptual roles, and which could be used to refer to the relevant objects or states of affairs in the external world.

The narrow intension of a belief (or thought) would determine which properties an object in the world would have to possess to be in the extension of that intension, i.e.,

the narrow intension determines extension conditions for the belief content in question.<sup>19</sup> This intensional method of denotation has been a recurring factor in this thesis, mainly because it seems to provide narrow content with referential properties. The principal reason for this is that the intensional reference of beliefs and thoughts provides a non-singular or descriptive reference, in contrast to a singular reference. A singular thought about an object would be about a single or particular object, whereas a non-singular thought about an object may be about several different objects, as long as those objects possess the relevant properties to satisfy the intension of the thought.

So, for instance, consider S and S\*'s narrow belief content that there is a glass of water on the table. If the term "water" is considered to be a singular term, then it could not be considered a term the meaning of which S and S\* shared. This is because in the actual situation, "water" would denote water or H<sub>2</sub>O, whilst in the counterfactual situation, "water" would denote twater or XYZ. So despite the identical descriptive properties the actual and counterfactual waters share, because the respective "water" terms are singular and because the respective waters are distinct natural kinds with distinct microstructures, they are considered to be distinct objects.

However, if the term "water" is considered to be a non-singular term, then it can be considered a term the meaning of which S and S\* share. In the actual and counterfactual situations, "water" would denote the transparent, odourless, etc. liquid in the local environment. So despite the actual and counterfactual waters having distinct microstructures (being distinct objects), because both waters have certain descriptive properties they satisfy the narrow intension of "water", this latter being something like "the transparent, colourless, potable, liquid in the local environment".



Therefore, the total narrow intension of S and S\*'s shared belief content that there is a glass of water on the table would have the referential property of picking out a situation in the world where there is a glass of liquid called water on a table in front of them.

### Narrow Content and Direct Reference

This specification of the referential properties that the narrow belief content shared by S and S\* has sounds reasonable enough. However, considering that the situation of the glass of water being on the table is right in front of S and S\* as they are perceiving it, it seems to me to involve a somewhat indirect method of denotation. From the subjective perspectives of S and S\*, they are presumably not having the occurrent thought that there is the transparent liquid called “water” in a glass in front of them, as they have their respective beliefs that there is a glass of water on the table. For instance, it could be imagined that S and S\* had just filled their respective glasses with water from the cold tap, before placing the filled glass on the table in front of them, i.e., they both have a shared belief that it is water in the glass in front of them in a manner that seems more direct than by a descriptive theory of reference. Is there a possibility that the narrow belief content shared by S and S\* could produce a more causally direct method of denotation, at least from their first-person perspectives? Perhaps we need to examine more closely the issue of singular and non-singular thoughts and terms, and to this end it may be useful to consider some of the history involved with this topic.<sup>20</sup>

To begin with, consider the following sentence:

(1) Dan Brown is the author of *The Da Vinci Code*.

At first glance, (1) seems to be a true identity statement, with the proper name “Dan Brown” singularly referring to Dan Brown and the definite description “the author of *The Da Vinci Code*” also singularly referring to Dan Brown, who *is* the author of *The Da Vinci Code*. Indeed, proper names and definite descriptions were considered to be *singular* terms, which had particular referents, and whose only function was to introduce their particular referents into the proposition that was to be expressed from a sentence which contained them. The upshot of this for (1), is that the proposition the sentence expresses would be equivalent to:

(2) Dan Brown is Dan Brown.

What has occurred in (2) is also an example of a well accepted principle that coreferring expressions can be substituted in a sentence without altering the truth value of that sentence. That is, the expression “the author of *The Da Vinci Code*” has been substituted with the coreferring expression “Dan Brown” and the truth value of the sentence has not altered, it is still true.

However, it was Frege (1892) who first brought attention to the fact that this substitutivity principle seems to be invalid when the sentence concerned is contained in a propositional attitude statement or that-clause. For example, someone could believe that (2) is true, and yet also believe that (1) is false. Indeed, Frege noticed that

whilst it could be said that (2) is trivially true, this is not the case with (1). If (1) is true, this seems to add to our knowledge, and so point to the fact that the singular expressions “Dan Brown” and “the author of *The Da Vinci Code*” have distinct meanings or contents, despite them being coreferential. Frege’s answer to this involved his famous distinction between the “sense” and “reference” of an expression. According to Frege, when a referring expression is contained in a that-clause, the expression not only “designates its reference”, but also “expresses its sense” (Frege, 1892, p.27), where the latter is understood to be the “mode of presentation” of the reference (Frege, 1892, p.24). The upshot of all this is that as far as Frege is concerned the propositions expressed by (1) and (2) when they are contained in that-clauses are different, because some of the expressions contained in the sentences contribute their senses to the proposition and not their referents. In particular, in (1) the expression “the author of *The Da Vinci Code*” is not contributing the referent Dan Brown to the proposition expressed, but the sense “the author of *The Da Vinci Code*”. The result of this is that when (1) is contained in a that-clause, the proposition expressed is that “Dan Brown is the author of *The Da Vinci Code*”, whilst when (2) is contained in a that-clause, the proposition expressed is that “Dan Brown is Dan Brown”.<sup>21</sup> This then is Frege’s reasoning for explaining the distinct meanings of the coreferring expressions “Dan Brown” and “the author of *The Da Vinci Code*” when they are contained in a that-clause.

Frege’s distinction between the sense and reference of an expression has of course been previously mentioned, particularly in relation to the conceptual role construal of narrow content and Loar’s use of the London/Londres, arthritis/arthrite etc. examples. However, how could one describe what has happened in the above explication of

Frege's distinction particularly in relation to the issue of singular and non-singular thoughts and terms? It seems that when (1) is outside a that-clause, then the proper name "Dan Brown" and the definite description "the author of *The Da Vinci Code*" are both functioning as singular terms and are coreferring to the same particular individual, i.e., Dan Brown. Using the terminology of Quine (1956), perhaps we could say that both referring terms are functioning in a transparent or de re way.

When (1) is contained in a that-clause the situation seems to change, particularly if the person having the belief that (1) is not sure whether Dan Brown is the author of *The Da Vinci Code*. With regards to the proper name "Dan Brown", I think that it is plausible to claim that the reference of this term would still be transparent and pick out Dan Brown. With regards to the sense of "Dan Brown", for the person having the belief that (1), this would not of course be "the author of *The Da Vinci Code*", as he or she is not sure that Dan Brown is the author of *The Da Vinci Code*. Perhaps it would be something like the *possible* author of *The Da Vinci Code*, or even just the author called Dan Brown.

However, what about the definite description "the author of *The Da Vinci Code*", what would be the sense and reference of this expression in a that-clause? It seems fairly clear that the sense of this expression is simply going to be the author of *The Da Vinci Code*, which is the mode of presentation or conception that the person having the belief that (1) has of the referent of that expression. In relation to the reference of the definite description "the author of *The Da Vinci Code*", things are a little more complicated. If one takes an objective third person perspective on the person's belief, it could still be claimed that the definite description has a singular reference which

denotes Dan Brown, even though the person having the belief may not be sure about this. However, it seems more plausible here that we are meant to adopt the subjective first-person perspective of the person when considering what the reference of the definite description is inside a that-clause. The result of this, if I am understanding things correctly, is that the reference of the definite description will now effectively be the sense of that definite description, i.e., the author of *The Da Vinci Code*.

Because the person having the belief that (1) does not know that Dan Brown is the author of *The Da Vinci Code*, then the definite description no longer has the singular reference of Dan Brown, but now has a non-singular reference which is basically whoever is the author of *The Da Vinci Code*, i.e., the definite description is now referentially opaque or *de dicto* in Quinean terminology.

This that-clause transformation of a singular term to a non-singular term shares the same features that were involved earlier with the narrow referring in situation (i).

When a term (or thought) is singular, it refers to a particular person/object, but when the term is transformed into non-singular mode, it now refers to whichever person/object in the world that possesses certain properties. Of course, one could still say that the non-singular term will pick out a *particular* person/object, as it may be a *unique* possession of properties that we are dealing with. However, this would still not be as directly referential as the functioning of a singular term, as the latter would refer to the same particular person/object in different possible worlds, whilst a non-singular term may refer to different particular persons/objects in different possible worlds, depending on which person/object had the relevant properties in that particular world.

If using a possible worlds semantics, an illustration of this would need to use a rigidifier such as “actual” or “actually” for the singular term but not for the non-singular term.<sup>22</sup> So, for example, with the singular version of “the author of *The Da Vinci Code*”, to rigidify the reference of this term one could use the statement “the actual author of *The Da Vinci Code*”. This would be a singular term as it would refer to Dan Brown in the actual world and also in all other possible worlds where Dan Brown existed, even if he had not been the author of *The Da Vinci Code* in some of those other worlds.

In the case of the non-singular version of “the author of *The Da Vinci Code*” this would not be rigidified with the use of “actual”, and so would refer to Dan Brown in the actual world, but may refer to someone else in another possible world if they were the author of *The Da Vinci Code* in that possible world, i.e., in another possible world it might be Jeffrey Archer who wrote *The Da Vinci Code*, and so the non-singular term would refer to him in that world, and not to Dan Brown. Indeed, this situation seems to put us in a bind, if it is a more “directly referential” narrow content that we are after. The reason for this is that it is the very “indirectly referential” nature of narrow content that enables the reference of that content to slide across possible worlds and so pick out different persons/objects in those worlds, i.e., it is the indirect referential properties of narrow content that enables it to be identified as narrow content in the standard Twin Earth-type scenarios.

Nevertheless, at around the same time as Frege, there was another philosophical theory in competition with Frege’s own, that attempted to answer the problem of the apparent invalidity of the substitutivity principle in sentences like (1) when they were

contained in that-clauses. This was Russell's (1905) famous and influential Theory of Descriptions. Russell's claim is that even though (1) looks like an identity statement, this is only a superficial appearance, at a deeper, logical level, it is no such thing.

Firstly, in order to explain this theory, let us assume that in (1) the proper name "Dan Brown" is a singular term which has a single, particular, referent (Salmon/Soames, 1988, p.4).<sup>23</sup> What Russell would now claim is that the definite description "the author of *The Da Vinci Code*" in (1) is not a singular term. He believed that the content or meaning of a definite description is not completely determined by its referent, and that it had an underlying structure composed of non-singular generalisations containing quantifying terms (Lycan, 2000, p.16). The upshot of this is that for Russell, sentence (1) would be an abbreviation for the following:

At least one person is the author of *The Da Vinci Code*

At most one person is the author of *The Da Vinci Code*

Whoever is the author of *The Da Vinci Code* is one and the same as Dan Brown.

The above can also be expressed using a variable as follows:

There is an individual  $x$  such that  $x$  and no one else is the author of *The Da Vinci Code*, and Dan Brown =  $x$ .<sup>24</sup>

Russell's Theory of Descriptions is a little complicated, but I think that the above is pretty much how Russell would have decomposed a sentence like (1) using his theory.

The question now is what bearing all of this has on the invalidity of the substitutivity principle when (1) is contained in a that-clause. The answer is that the above does explain the apparent invalidity of the substitutivity principle, because the meaning or content that the proper name “Dan Brown” and the definite description “the author of *The Da Vinci Code*” contribute to the proposition expressed by a sentence containing them is not the same. The proper name contributes its single referent or bearer Dan Brown, whereas, the definite description contributes a general relational property to the effect that there is an individual  $x$  such that  $x$  and no one else is the author of *The Da Vinci Code*. In the case of (1), Russell’s theory results in this general relational property being attributed to Dan Brown, but this is a contingent attribution, which might have been different in different circumstances. The upshot of this is that according to the Theory of Descriptions, “Dan Brown” and “the author of *The Da Vinci Code*” are not really coreferential terms at all, and so may not be intersubstitutable *salva veritate* in a sentence which contains them. Another way to put this is that there may well be different propositions expressed from a sentence which interchanged the terms “Dan Brown” and “the author of *The Da Vinci Code*”.

With regards to explaining the apparent invalidity of the substitutivity principle, there does seem to be some similarity between Russell’s theory and Frege’s theory. In both theories we have seen an apparently singular term, i.e., “the author of *The Da Vinci Code*”, be transformed into a non-singular term whose reference is not so direct and particular as that of the proper name “Dan Brown”, which in both cases has remained a singular term. It seems that whenever a definite description is used to denote an object or person in the world, particularly when that description is contained in a that-clause, the chances are that it is not going to be a singular or direct reference.



This point can be further illustrated by considering Russell's strategy of extending his Theory of Descriptions to apply to proper names. This involved Russell claiming that proper names contributed more meaning to propositions than just their bearers, the latter of which was the view usually attributed to J.S. Mill (1843). Indeed, Russell claimed that proper names were actually abbreviations of definite descriptions, so that, for example, the proper name "Plato" was equivalent to the definite description "the pupil of Socrates and the teacher of Aristotle" or something like that.

However, as Kripke pointed out, the upshot of this is that the equivalent definite description for a proper name does not directly fix its referent or bearer, but rather, it once again exemplifies the non-singular referential properties that are generally involved with definite descriptions. For example, consider the above definite description associated with the proper name Plato. Roughly, Kripke's argument is that it is possible that someone else might have been the pupil of Socrates and teacher of Aristotle, say it was Ronald Spriggs of 23 Colosseum Way, Athens; whereas, Plato was just a humble toilet cleaner. The intuition that Kripke wanted us to have was that in contrast to a descriptive theory of reference, a causal theory of reference would fix its bearer as the particular individual who was Plato, so that in a possible world where Ronald Spriggs satisfied the relevant (Plato) definite description, the proper name "Plato" would still pick out the same particular individual, who in this case was a toilet cleaner.

This situation with proper names seems similar to the situation with natural kind terms. A natural kind term like "water" behaves like a proper name, in that its

reference is fixed by the causal connection it has to the particular stuff that is called “water” in that world. In the actual situation, “water” gets fixed to H<sub>2</sub>O, whereas in a counterfactual situation, where say, H<sub>2</sub>O is a black tarry substance and the watery-stuff is composed of XYZ, a causal theorist of reference would probably argue that the natural kind term “water” should correctly be used to refer to the black tarry substance, rather than the watery-stuff in that world. That is, using Kripke’s terminology, a singular term could be understood as a rigid designator across possible worlds, whereas a non-singular term could be understood as a non-rigid designator across possible worlds. The result of this is that a definite description would probably be considered a non-rigid designator, as we have seen previously that its reference does slide across possible worlds.

At this point, it might be worthwhile to pause for a moment, and consider where we have got to in this discussion about singular and non-singular thoughts/terms. The reason for the discussion was that it seemed to me that the narrow representational content shared by S and S\*’s respective beliefs that there is a glass of water on the table did not have the directness or immediacy that it perhaps should have, given that the situation they were having their shared belief about was right in front of them. It was deemed that the term “water” in the beliefs would have to be a non-singular descriptive term rather than a singular term, in order that it could refer to both of the distinct natural kinds called “water” in the actual and counterfactual situations. Now, from the above discussions of Frege and Russell in relation to singular and non-singular thoughts/terms, it does seem that when descriptive methods of denotation are involved, then the relevant thoughts/terms are going to become less direct and particular and more indirect and general. So this seems to back up the earlier

conclusion that the term “water” in S and S\*’s shared belief would have to possess descriptive non-singular referential properties. Therefore, is there any possibility of producing a more causally direct, perhaps singular, reference, for the narrow “water” term that could still enable it to refer to either H<sub>2</sub>O or XYZ? It seems to me that there may still be a possibility, but it will involve us going back to some of Russell’s ideas on reference and knowledge.

### Russell’s Notion of Acquaintance

I may as well let the cat out of the bag, so to speak, and reveal that the possibility I am referring to involves Russell’s well-known notion of knowledge by acquaintance.

However, I am going to approach this notion under the influence of McDowell (1986), who is also investigating the issues involved with singular and non-singular thoughts/terms, particularly in relation to Russell and Frege. Therefore, let us return to Russell’s Theory of Descriptions, which we have seen is used by Russell to explain the apparent failure of the substitutivity principle with coreferring terms contained in a that-clause. However, the Theory of Descriptions was viewed by Russell as an apparatus which could solve several more apparent “puzzles” involved with meaning and reference (Russell, 1905, p.47). The particular “puzzle” I have in mind is the one which involves the apparent reference of singular terms to persons/objects which do not actually exist. To illustrate this, consider the following sentence:

(3) The present Swansea University Philosophy Department has 20 members of staff.

The above seems to have a meaning which is readily understandable. However, the sad and shameful truth of the matter is that Swansea University doesn't have a Philosophy Department anymore (but it does have a great Business School...), so the apparently singular term "Swansea University Philosophy Department" is referring to a nonexistent entity. Now, the problem this creates is that (according to McDowell), for Russell, the term "singular" in singular term/thought basically means "object-dependent", which has the result that sentences containing singular terms can only express singular propositions if the referent of the singular term does exist (McDowell, 1986, p.137). Therefore, if for Russell the meaning of a sentence containing a singular term is given by the singular proposition which the sentence expresses, then if the relevant singular term in a sentence does not have a referent, the sentence would not express a singular proposition and so not have a meaning; or as Russell puts it, the sentence "...ought to be nonsense" (Russell, 1905, p.46).<sup>25</sup>

Therefore, because the singular definite description "the present Swansea University Philosophy Department" in (3) does not have a reference, then on Russell's reading it should not have a meaning and just be nonsense. However, this is the "puzzle", because (3) does seem to have an understandable meaning and could be said to be false given the circumstances. In order to explain this, Russell brings in his Theory of Descriptions, so that (3) can be decomposed as follows:

There is at least one present Swansea University Philosophy Department

There is at most one present Swansea University Philosophy Department

The present Swansea University Philosophy Department has 20 members of staff.

Therefore, according to Russell, the definite description “the present Swansea University Philosophy Department” is not a singular term, or equivalently, it is not a genuinely referring expression. As before, the meaning of the definite description does not depend solely on its referent, but also seems to depend on an underlying structure consisting of several non-singular sentences. This explains the puzzle because it allows (3) to express a proposition whether or not there is an existing referent which is denoted by the non-singular or descriptive sentences given in the decomposition (McDowell, 1986, p.137). Therefore, it can be seen that (3) appears to be false because the first decomposed sentence is false, i.e., it is false that there is at least one present Swansea University Philosophy Department.<sup>26</sup>

Now, according to McDowell, the effect an example like the above had on Russell’s thinking in relation to singular terms and singular propositions resulted in (what McDowell calls) “Russell’s restriction”, which is that:

...we can entertain and express singular propositions only where there cannot be illusions as to the existence of an object of the appropriate kind: only about features of sense-data or items present to us with similar immediacy in memory, and (when Russell recognized them as objects) our selves (McDowell, 1986, p.138).

Therefore, it seems that at this particular time in his philosophical life, when Russell talked of singular terms and propositions, the particular entities that were the objects of the singular terms and propositions were not external physical objects, but internal properties of experienced sense-data. As McDowell points out (McDowell, 1986, pp.145-146), this is a very Cartesian outlook on the relationship between thought and

reality, where a kind of first-person authority results in a subjective certainty of the contents of one's own mind, and where all the uncertainty and errors are placed firmly in the realm of the external world. Indeed, Russell (1910/11) seems to emphasise this "internalist" outlook on the relationship between thought and reality with his apparatus of knowledge by acquaintance and knowledge by description. He declares that:

We began by distinguishing two sorts of knowledge of objects, namely, knowledge by *acquaintance* and knowledge by *description*. Of these it is only the former that brings the object itself before the mind. We have acquaintance with sense-data...but not with physical objects or other minds...Our knowledge of physical objects and of other minds is only knowledge by description, the descriptions involved being usually such as involve sense-data. All propositions intelligible to us, whether or not they primarily concern things only known to us by description, are composed wholly of constituents with which we are acquainted (Russell, 1910/11, p.31).

The above quote from Russell does appear to have a contradictory air about it, with him claiming that it is only acquaintance that brings the "object itself" before the mind, but then claiming that while we do have acquaintance with sense data, we do not have acquaintance with "physical objects". My own reading of this is that when Russell claims that only acquaintance brings the "object itself" before the mind, what he means is that it is the *representation* of the physical object itself, which is constituted by sense data, that is before the mind. If this is accepted, then it can be seen that the foundations of Russell's epistemology once more comes back to the properties of experienced sense-data, with knowledge of physical objects only coming indirectly through the direct knowledge we have of our own sense-data.

Now, McDowell does not agree with Russell's epistemology, indeed, it could be claimed that the crux of his 1986 article is that he argues that singular thoughts and

propositions can be entertained and expressed outside of Russell's restriction and would specifically involve external objects that are being perceptually and demonstratively referred to by the subject (McDowell, 1986, p.140). Moreover, I have the feeling that McDowell would probably think that I have been under some sort of Cartesian spell in this chapter, with my talk of narrow representations, subjective points of view, etc. Nevertheless, it seems to me that Russell's talk of acquaintance with sense-data does have some similarity with the earlier talk of narrow representational content, which involved a subject referring to an internal representation with a mental symbol.<sup>27</sup> One could perhaps describe this situation as the subject being causally acquainted with a particular internal representation, the latter of which is of course the cognitively-processed result of certain proximal stimulations on the subject's sensory receptors. The notion of a sense-datum theory is notoriously vague, however, as Gallois (1998) points out, it seems that the following claims can be made:

[a] Every perceptual experience involves awareness of a sense-datum.

[b] At least for a certain range of characteristics, it is impossible for a sense-datum to appear to have a characteristic that it lacks (Gallois, 1998, p.695).

With [a], it seems clear that a sense-datum theory involves a representational theory of perception, where the experienced sense-data are representing the external world in a certain way, i.e., the subject is not immediately perceiving the external world but only mediately perceiving it, via the sense-data.<sup>28</sup> Once again, this sounds very much like the situation that has been described in this chapter concerning narrow representational content, as the internal representations are mediating the subject's contact with the external world. As for [b], this claim ties-in with Russell's

restriction, where we can have singular thoughts or propositions about the properties of sense-data, as there is no possibility of illusion with these properties. In the case of narrow representational properties, perhaps we could say that if a certain substance called “water” is represented as being a colourless and transparent liquid, then the internal representation necessarily has these properties, despite what extra-mental properties “water” may actually have.

### Acquaintance and the Possibility of “Direct” Reference?

However, it is all very well to point out the similarities between the notions of sense-data and internal representations, but how does this help us get a more causally-direct method of narrow or shared reference in situation (i), where S and S\* are perceiving and believing that there is a glass of water on the table? What I have in mind is the following. The narrow representational content that is shared between S and S\* is postulated to be a virtual reality of Situation (i), i.e., S and S\* both have identical internal representations of a glass of water on a table which appear to be located and extended in space. However, the internal representations are only neurophysically constructed from the proximal stimulations on S and S\*'s sense organs, so the distal objects (glass, water/twin-water, table) which are presumed to be causally responsible for producing the relevant proximal stimulations are not included in the virtual construction of situation (i).

What can perhaps be said is that both S and S\* are acquainted with their respective internal representations and, following Russell and his sense-data, they can entertain



and express singular propositions about those internal representations. I would claim a singular proposition can be expressed about the shared “Water” symbols and the [transparent liquid] representations they internally refer to, even though the respective distal waters may have different microstructures. This is because at the level of reality that the representations are operating at, i.e., the level of (in this situation) the phenomenal, visible, properties of water, then both the actual and counterfactual internal [transparent liquid] representations would count as the same singular object. I think I am able to make this claim as a component (that is provided by the narrow belief) of the internal representational content of both S and S\* is that there is a transparent liquid called water in the respective glasses in front of them. That is, with both S and S\*, they are representing a situation where it seems to them that there is water in the glasses. If the liquid in the glasses is not water, then their shared narrow representations will not be correct and their resultant beliefs will be false. However, this latter point can still be accepted, without it affecting the content that S and S\*’s internal representations are claimed to have.

Therefore, if we now imagine the narrow virtual reality of situation (i), but this time, also imagine superimposing the narrow representations of the objects involved over the distal objects which are causally responsible for producing those representations. It seems to me that the intuitive result of this superimposition would be that we had gone from a shared *virtual* situation (i) to a shared *veridical* situation (i). It would be a shared or narrow veridical situation (i), because the narrow internal representation of situation (i) could be exactly superimposed over either the actual or counterfactual veridical situation (i).

If all of this is allowed, the upshot would be that the internal representations of S and S\* would have been spread-out over the causally-relevant distal objects of a narrow situation (i). Therefore, as S and S\* were directly acquainted with their narrow representational content, because that narrow content has now been effectively dragged out into the external world by the superimposition, perhaps it could be said that they are able to narrowly refer to situation (i) in a more causally-direct manner than they had previously. Moreover, it might be feasible to claim that with the narrow veridical situation (i), (which I am calling narrow because it involves only the identical representational properties of the distal objects involved), both S and S\*'s "water" terms/thoughts could be singular in nature, because it would be at the level of the visible or macro-properties of the distal water, that the singular denoting would be taking place.

### **8.3 Narrow Content and Psychological Explanation**

The result of the above examination of situation (i) also has the related motivation that it seems plausible that the possible behaviour of S and S\* in that situation can have a psychological explanation that subsumes the both of them. The narrow content that is shared between S and S\* could be specified by the sentence "S/S\* believes that there is a glass of water on the table", where this sentence could be understood to be the narrow or subjective de dicto version of the belief. This specification would be intended to convey the point that the way S and S\* internally represent situation (i) is identical, i.e., they both internally represent water as a transparent, odourless, potable, etc., liquid.

The upshot of this is that if S and S\* proceed to drink their respective glasses of water until the glasses are empty, then the explanation of their identical behaviour can be subsumed under a psychological generalisation, something like: S and S\* were both thirsty, they believed that water quenches thirst, they further believed that there was a glass of water on the table, they desired to quench their thirsts, so they drank the water. On an externalist reading, the twins' behaviour could not be so subsumed under the same psychological explanation, because for the externalists the explanation would involve different beliefs, i.e., S's beliefs would be about water (H<sub>2</sub>O) whereas S\*'s beliefs would be about twater (XYZ). Indeed, externalists would also claim that S and S\*'s behaviour was not identical either, with S drinking a glass of water and S\* drinking a glass of twater.

An example which illustrates this kind of thinking would be Macdonald (1998), who tries to provide an argument to show that a narrow construal of twins' actions in a Twin Earth-type scenario is dependent on a background of "wide-act individuation...[where] wide acts are only made intelligible by states with wide content" (Macdonald, 1998, p.290). Macdonald imagines a Twin-Earth situation where Sue washes with water whilst her twin washes with twater. She admits that the actions of Sue and her twin could be given a credible narrow-act explanation which does not advert to wide content, but claims that:

...the intelligibility of *what* one is doing, narrowly construed as a successful activity, takes place against the background of assumptions about what it is *appropriate* to do it *with*. One does not make intelligible Sue's activity of washing every day just by mentioning her desire to make herself clean and her belief that by washing she will make herself clean. Her activity simply does not count as an activity of washing if she does it with mud, or Coca-Cola, or

with tar. And here I mean: *successful* activity. Her movements may be the movements of someone who takes herself to be washing. But movements are not actions; and their classification as actions, even narrowly construed, depends on what the appropriate objects are towards which they are directed (Macdonald, 1998, p.291).

This sounds reasonable, but as Macdonald herself realises, it is open to the supporter of narrow-act individuation to point out that in the case of Sue and her twin, if their successful actions of washing need to be directed towards the appropriate objects in their environment, then the water and twater can be narrowly construed as the transparent, odourless, colourless liquid, that is appropriate in Sue and her twin's respective environments. The upshot of this is that there is still no need of any wide-act individuation to explain Sue and her twin's successful washing-actions. What Macdonald needs to do to bring in the wide-act individuation is to show how water qua H<sub>2</sub>O is the appropriate object for successful washing with Sue and (t)water qua XYZ is the appropriate object for successful washing with Sue's twin. She attempts to do this by going counterfactual, i.e., she imagines Sue transported to her twin's context and tries to argue that it is not appropriate for Sue to use twater to wash successfully, as it is only with water qua H<sub>2</sub>O in her actual context that she can wash successfully (Macdonald, 1998, pp.293-296).

It can be seen here that Macdonald is presupposing the Kripke/Putnam line on the individuation of natural kinds, which is basically a causal-historical one. This is a slightly surprising move by Macdonald, as earlier in her paper she had declared that "...nothing in the argument to follow requires commitment to any doctrine about natural kinds" (Macdonald, 1998, p.284). Be this as it may, it is still open to the supporter of narrow-act individuation to claim that the intuition they have is that in the counterfactual situation twater is the appropriate object that enables Sue to wash

successfully, as it is twater that is the transparent, odourless, colourless liquid that people wash with in that context. Macdonald's answer to this is to adopt Millikan's (1984, 1993) view of a biological function and to basically claim that there is no motivation to give water/twater a more general (and narrow) specification of being a transparent, odourless, etc., liquid; as she puts it "It all depends on the organism and what is in its actual environment" (Macdonald, 1998, p.n294).

However, it seems to me that a more general specification is only unmotivated if Millikan's view of a biological function is accepted in describing the successful washing actions of Sue in the counterfactual situation. If Millikan's view is not accepted, there then appears a very large motivation for giving a more general (and narrow) description to water/twater, namely, that it would allow the successful washing actions of Sue and her twin to be subsumed under the same psychological generalisation that explained their respective actions.

Therefore, even with a counterfactual situation of S being in S\*'s context (or vice versa), it seems far more plausible that their successful thirst-quenching actions can be subsumed under the same psychological explanation, rather than the externalist alternative. The reason for this is that a psychological explanation of behaviour (and action) surely depends on how the subject conceives or represents things to be, which in the case of S and S\* is identical, which therefore means that an explanation of their behaviour must take into account this identity. It might be an interesting thing to learn about the world that the respective waters in the actual and counterfactual situations were different substances, but it seems to me that this point is irrelevant when giving psychological explanations of S and S\*'s behaviour. If the way things

seem to S and S\* are the same, then even though there may be differences in their respective physical environments, the psychological explanation of their behaviour, if it is to accurately take into account the subjective first-person viewpoints of S and S\*, must also take into account the sameness of the seemings.<sup>29</sup>

Therefore, this is my (hopefully) adequate and plausible formulation of a narrow content, which has a mainly representational character, but has also included elements from the conceptual role and indexical construals of narrow content. Moreover, I believe that the resultant notion of narrow representational content is capable of satisfying all three conditions of adequacy for narrow content, something which the previous three construals were unable to do. It satisfies condition (1), positing a narrow belief content that is internal to, or “in the head” of, both duplicates in a Twin Earth-type scenario. It does this by making the narrow representational content locally supervenient on the neurophysiology of the duplicates. This was something that I thought Dennett didn’t do with his notional worlds, with the narrow belief content being attributed in an instrumental manner, based mainly on the subject’s behavioural dispositions. My narrow content also satisfies condition (2), in that it is capable of being given a plausible specification using that-clauses. This was something that Loar’s original conception of psychological content was incapable of. Finally, my narrow content satisfies condition (3), in that it provides a plausible psychological generalisation that subsumes the behaviour of both duplicates in a Twin Earth-type scenario. This was something that I thought Fodor’s mapping theory narrow content did not do, as it was not the shared narrow content itself that constituted the psychological generalisation, but the shared wide content in the same context. Moreover, it also seems plausible that my narrow content construal could

deal effectively with so-called Frege cases. After all, Frege cases usually involve a subject's different conceptions of the same object, a factor which a representational narrow content would be expected to deal with, as it tries to capture how the subject is representing, or conceiving of, things in the world.

It also seems to me that the above formulation would work with most standard Twin Earth situations involving natural kinds,<sup>30</sup> at least given the following two caveats.

(1) That the situations involve the representational or perceptual properties of the natural kinds in question. If this is the case, then the way that twins internally represent the natural kinds in question should produce some sort of narrow content that is shared between them. (2) That the narrow beliefs involved make specific mention of the natural kind in question, e.g., water, gold, aluminium, etc. The reason for this is that this enables the mental symbols the subjects use for denoting the distal natural kinds in question to also denote the internal representations of those natural kinds. This in turn enables the narrow beliefs to have narrow truth conditions, which puts a constraint on which phenomenologically identical external world referents can be correctly and truthfully contained in the extension of the mental symbol for the natural kind.

## Notes to Chapter 8

<sup>1</sup> In what follows, the locutions "perception", "perceptual experience", "perceptual content", etc. will be frequently mentioned. However, these locutions should be understood as meaning "*visual* perception", "*visual* perceptual experience", "*visual* perceptual content", etc.

<sup>2</sup> The upshot of this of course is that philosophers who support different theories of perception may not be too impressed or convinced about what is to follow. Rather ironically, I have the feeling that Putnam himself may be one of these. In his 1999 book, he espouses a so-called naïve or direct realist approach to perception. The basic thrust of this approach is that when we perceive objects in the external world, it is a direct or unmediated perception. That is, there are considered to be no mediating internal representational states involved with the perception of distal objects by the subject. Indeed, Putnam declares that "I believe this idea [of internal representations mediating perception of the external world] is responsible for the central complex of intractable "problems" and unworkable positions that has bedevilled philosophy since Descartes" (1999, p.128).

The main support for the idea of mediating internal representations comes from the intuitions that are pumped from comparing the perceptual experience of a veridical perception of an object and the perceptual experience of an hallucination of that same object. A powerful intuition that comes from this situation is that from the subject's viewpoint, the perceptual experience of the veridical situation would be indistinguishable from the perceptual experience of the nonveridical situation, i.e., it would seem that the veridical perceptual state is type-identical to the nonveridical perceptual state. But as there is no distal object in the nonveridical case, this seems to show that our perceptions are mediated by internal perceptual or representational states of some kind. For example, imagine that a person has a veridical perception of a red garden gnome and the same person also has an hallucination of the same red garden gnome, where the latter experience is understood to be qualitatively identical to the former experience. The intuitive upshot of this is that there is an internal perceptual state that is shared between both the veridical and nonveridical cases, which can be described as something like "There appears to be a red garden gnome in front of me". However, it is this assumption that Putnam questions. He takes a "disjunctive" view of the above situation, following other direct realists such as McDowell (1982, 1994) and also J.L. Austin and William James (Putnam, 1999, p.152). According to the disjunctive view, there is nothing in common between the veridical and nonveridical cases, there is only the disjunctive situation of one either seeing a red garden gnome if the experience is veridical or seeming to see a red garden gnome if the experience is nonveridical. That is, with the veridical experience, an object (the garden gnome) actually has the property of red, whereas, with the nonveridical experience, the property of red is only in the experience itself, it has not been attributed to an actual object (see 1999, pp.128-129; 151-154).

However, it seems to me that this disjunctive strategy is yet another example of a theme that has recurred throughout this thesis; namely, it is an argument that tries to support an externalist position by adopting an objective third-person perspective on a situation, rather than a subjective first-person perspective. With the above example, the objective perspective (or perhaps, the meta-perceptual position) is being assumed, as we know for sure that one case is veridical and the other is nonveridical. However, if we adopt the subjective perspective of the person having the experiences in both cases, he/she cannot tell which case is the veridical one and which is not. The reason for this is of course that both experiences are qualitatively identical, i.e., in each case it seems as if the person attributes the property red to an actual object, the garden gnome (Indeed, Putnam (speaking on behalf of McDowell) does acknowledge that there is a certain content that both the veridical and nonveridical experiences share (1999, p.154)). Perhaps we can say here, following Comesana (2005), that the veridical and nonveridical perceptual experiences would enable the subject to have a shared justified perceptual belief that there is a red garden-gnome in front of him/her, even though it is only with the veridical experience that the subject has perceptual *knowledge* that there is a red garden-gnome in front of him/her (2005, p.382). Therefore, if we are going to stay faithful to the first-person perspective of the subject and the notion of narrow content, I do not find the disjunctive theory of appearances very persuasive, as it seems to downgrade the importance of the subjective viewpoint.

Moreover, as Putnam denies that there is any internal representational state involved with visual perception, it would be interesting to see how he explains the subject "seeming to see a red garden gnome" in the above nonveridical case, without adverting to the idea of internal representations. That is, there is no distal red garden gnome in the nonveridical case, so where does the appearance of one come from? It seems to me that the only plausible answer to this is that it is an internal neurophysically produced appearance, which then leads to the following puzzling situation. Are we to believe that even though the same parts of the subject's brain and sensory receptors are active in the nonveridical case and veridical case, which in the nonveridical case produces the qualitatively identical appearance of a red garden gnome, in the veridical case there are no internal representational components involved in the seeing of the red garden gnome? It seems to me that to answer yes to this is unintuitive and not very persuasive.

<sup>3</sup> The accusation could again be made here that what I am doing is presupposing that the narrow content between twins is identical. However, as before, I would reject this charge. I am only presupposing that how the world seems to twins is identical. It is a further step to actually specify whether there is some narrow semantic content that the beliefs of the twins share.

<sup>4</sup> It should be noted here that the externalist might also accept this diagnosis. However, from my readings of the literature, it seems that most externalists are dubious about the idea of a narrow content, e.g., Baker (1987a, 1987b), Stalnaker (1990), McCulloch (1995).

<sup>5</sup> I suppose that what I am suggesting here could be described as finding out how much of the subject's environment comes with the subject when he or she is removed from it mid-experience and mid-believing.



<sup>6</sup> The useful expressions “perceptual processing” and “neural representations” are taken from Velmans (1998).

<sup>7</sup> I wonder if this idea of a subjective virtual reality is similar to Aristotle’s views on perception? Aristotle believed that a physical object was composed of form and matter, but that a perception of a physical object only had the form, which shaped the perceiver’s mind accordingly. Perhaps the idea of the object’s form shaping the perceiver’s mind is similar to the visual phenomenology of the subject’s perception of an object, constructed from the proximal stimulations in. Moreover, the resultant narrow virtual reality that is produced from this visual experience would of course not include the physical matter that is deemed to be “under” the surface appearance of the object. See Cummins, 1989, pp.2-4 and Crane, 1998, p.817, for brief but interesting accounts of Aristotle’s views on perception and intentionality.

<sup>8</sup> This denotation of internal representations using mental symbols that I have in mind is similar to Perlis’s (1991) notion of internal reference, where the subject intends to use a certain mental symbol to refer to an internal representation. One difference is that I wouldn’t claim that S/S\* are “intending” to refer to their internal representations, as this seems to imply that it is a conscious process of internal referring that occurs. However, I’ll follow Perlis in using “” to identify the mental symbols and [] to identify the internal representations, so that, for example, the mental symbol “Table” denotes the internal representation [Table].

<sup>9</sup> Moreover, this claim of shared narrow content between S and S\* could also be supported by a narrow conceptual role reading of this situation. That is, both S and S\* make what would be deemed the same appropriate and correct (instantaneous and unconscious) use of certain mental symbols to refer to certain internal representations.

<sup>10</sup> This situation also provides a new angle on what is usually called the transparency of perceptual experience. The claim is made that, for example, if the subject is having a visual experience of a glass of water on a table, and he/she then introspects the experience to try to find something intrinsic or internal to that experience, then nothing can be found. All that the subject finds is the intentional content of the experience, i.e., the glass of water on a table, which are considered to be external to the subject. This result could then be used to support the claim that there are no internal components involved with the visual perception of external objects, or the claim that Tye (1995) makes, where he declares that “The lesson of the problem of transparency is that *phenomenology ain’t in the head*” (1995, p.151). However, if what I have claimed previously is accepted, just because a visual experience only seems to involve objects which are external to the subject, this does not entail that those objects are actually external to the subject, perhaps they are internally constructed representations of a distal state of affairs, i.e. a narrow virtual reality, as with the situation of S and S\* when attached to the alien supercomputer. Just because the content of a visual experience is intentional, this doesn’t mean that the content can’t be determined by the subject’s internal physical properties. This also means, contra Tye, that phenomenology may well be in the head.

<sup>11</sup> This is perhaps just as well, as so-called causal covariance theories of meaning or representation, e.g. Fodor (1987, 1990), seem to have a major problem. In such theories, a claim such as the following is usually made: x means or represents y due to instantiations of x nomically covarying with instantiations of y, i.e., it is a law that every time a y is instantiated, an x is also instantiated. However, the problem for this theory of mental representation is what Fodor calls the disjunction problem, which basically involves situations when a mental symbol x is instantiated, but not due to the instantiation of its lawful distal covariant y, but due to the instantiation of, say, a distal z. The upshot of this is that the mental symbol x now has the disjunctive representational content of “y or z”. The further upshot of this is that for causal covariance theories, misrepresentation can never occur. The reason for this is that whatever different distal object/property instantiations cause the instantiation of the same mental symbol token, these different objects/properties just become disjuncts of the representational content of the mental symbol, so there is never the possibility of the symbol misrepresenting an external world object or property.

<sup>12</sup> Therefore, it could be claimed here that S and S\* have narrow belief tokens of the same de dicto semantic type. From my reading of Taylor (1989a, 1989b), I think that he would probably agree with the above claim. However, Taylor takes the view that in his so-called Fraternal Twin Earth scenarios, despite fraternal twins being functionally type-identical, they could still have type-distinct de dicto beliefs (1989a, p.364). Taylor’s target here is the claim of narrow content functionalism that if twins have beliefs which are functionally type-identical, then those beliefs will also be of identical de dicto semantic type (1989a, pp.355-357). Briefly, and using the above situation (i), Taylor’s idea would be to claim that in the counterfactual scenario of S\*, the “water” in the glass would not be a liquid qua liquid, as this is understood on Earth, or indeed, as usually understood on Twin Earth. It would be a

substance that is phenomenally identical to water, but would be "...made up of very fine, but not microscopic, particulate matter" (1989a, p.364), which are "...like extremely fine grains of sand" (1989a, p.n372). The upshot of this for Taylor would be that S/S\*'s narrow belief tokens that there is a glass of water on the table would be of distinct (and so not shared) de dicto semantic types. Taylor's conclusion, if I understand him correctly, comes from his definition of de dicto belief types, which involve descriptive constraints on what content they have. Therefore, because the counterfactual water has a different descriptive constraint to the actual water, i.e., the former has a description like "it is made up of very fine particles like sand", then a belief token containing the counterfactual "water" term would be of a different de dicto semantic type to a belief token containing the actual "water" term.

However, I don't find Taylor's argument very compelling, particularly as the identity of the qualitative or phenomenal properties of the respective waters seem to count for nothing as far as he is concerned. He talks of using "qualitative concepts" to narrowly individuate de dicto beliefs but is dismissive of this idea, seeming to believe that this way eventually leads to "...phenomenalism or a crude sense datum theory" (1989a, p.370). Briefly, I don't think there's anything "crude" about the idea of a subjective de dicto belief type which is individuated by the qualitative (as well as conceptual) properties of experience. Even if counterfactual "water" is considered to be composed of fine sand-like particles, if this difference does not impinge on the phenomenal and conceptual experience of S\*, then it seems to me that the properties of this experience can still be used in the production of type-identical subjective de dicto water beliefs which are shared between S and S\*. That is, in the terminology of Taylor, S and S\*'s subjective de dicto water beliefs can still be type-identical because the respective "water" terms can still have the same descriptive constraints, e.g., local transparent liquid.

<sup>13</sup> In the above I have deliberately used Nagel's (1974) phraseology which he used in his discussion about the problems of consciousness.

<sup>14</sup> A similar objection is also made by Segal, 2000, pp.114-115, where he is discussing the addition of a meta-linguistic component to the descriptive character of a natural kind concept, whose function is to help determine the narrow cognitive content of that concept.

<sup>15</sup> Interestingly, this seems to entail that whichever liquid is legitimately contained in the extension of "Water", this liquid will have to have all the representational and functional properties that water has in the actual world. For example, in a counterfactual world where 7-UP is conventionally called water, that nice taste of lemon would have to go, as water is considered to be tasteless, also the bubbles will have to go, as conventional water is naturally non-carbonated, etc. Therefore, whatever liquid is chosen to be conventionally called water, the narrow representational/functional properties of the internal [transparent liquid] can be used to mould that liquid into the conceptual shape of actual water.

<sup>16</sup> The following précis of Baker's thought experiment comes mainly from (1987a, p.207). It is here that Baker acknowledges that her thought experiment is similar to ones put forward by Tyler Burge and Stephen Stich. However, it seems to me that Baker's thought experiment bears the closest similarity to Putnam's (1983) "Grug" thought experiment. Briefly, Putnam imagines a country called Ruritania where a metal called grug is used to make pots and pans. However, it turns out that the north and the south of the country have different dialects, which results in "grug" meaning silver in the north and "grug" meaning aluminum in the south (1983, p.529). The question is now whether the belief "Pots and pans are made of grug" has the same meaning when uttered by two doppelgangers, one in the north of the country, the other in the south. Putnam concludes that the northern grug has a different meaning to the southern grug, due to the different extensions for the northern and southern grugs, i.e., the northern grug denotes silver and the southern grug denotes aluminum (1983, p.530). Indeed, Putnam once again claims that "Meanings aren't in the head", due to the fact that there is a difference of meaning between the north grug and south grug, even though the doppelgangers have identical mental representations or conceptions of grug (1983, p.530). However, could it be that Putnam should only claim that *some* meanings aren't in the head? That is, can the identical mental representations or conceptions of grug also provide an identical meaning for that term? Putnam himself doubts this, at least in relation to a computational psychology. However, it is this question that will be addressed as we examine Baker's thought experiment, which I chose to concentrate on rather than Putnam's thought experiment, as the former bore more resemblance to the 7-UP objection to narrow truth conditions just considered.

<sup>17</sup> Having said this, it is interesting to note (which my supervisor kindly noted) that if the narrow construal for vodka included something like the following, "vodka is a clear liquid and tastes like...", it seems plausible to come to the conclusion that the narrow belief contents of the two girls could still differ, due to the fact that vodka and gin don't taste the same. This is a tricky issue, but perhaps the following can be said. As far as Baker's thought experiment is concerned, the two girls are considered to be molecule for molecule duplicates, which entails, if one accepts my earlier mentioned

presupposition about the supervenience of the subject's internal representational properties on neurophysiological properties, that molecular duplicates will have identical representations of how things seem to be. However, if it seems to one girl that "vodka" tastes like surgical spirit and to the other girl that "vodka" tastes like rum, then presumably the girls would no longer be molecular duplicates and so may well have different narrow contents due to their different internal representational properties.

<sup>18</sup> Of course, it is feasible to consider conceptual roles themselves as types of representational states, but I will continue to keep them separate for the purpose of specifying types of narrow content.

<sup>19</sup> See Searle (1983), Horowitz (1996) and Jackson (1998) for discussions on this topic.

<sup>20</sup> In the following discussion of the work of Frege and Russell, I have been particularly influenced by Blackburn (1984), Forbes (1998), Gallois (1998), Lycan (2000), McDowell (1986), Salmon/Soames (1988), Taylor (1998).

<sup>21</sup> In a similar way to the different propositions expressed by (1) and (2), it might be said that (1) and (2) considered as complete sentences would also have different senses. Frege was of the opinion that a complete sentence in a that-clause also had a sense, which in this case would be the thought that the sentence contained (1892, p.28). This sounds complicated, but it turns out to be quite straightforward. For instance, the thought contained in sentence (1) is just the thought that Dan Brown is the author of *The Da Vinci Code*, whilst the thought contained in sentence (2) is just the thought that Dan Brown is Dan Brown. From this it can be seen that as the thoughts contained in (1) and (2) are different, then the senses of (1) and (2) are also different. The upshot of this is that when (1) and (2) are contained in that-clauses, one reason for them not to be considered equivalent is that they express different senses. However, given the above, it seems plausible to assume that for Frege, the thought/sense contained in a sentence was equivalent to the proposition the sentence expressed.

<sup>22</sup> See Davies/Humberstone (1980) about the use of "actually" as a rigidifier; see also Braddon-Mitchell/Jackson (1996) for a clear discussion of these issues.

<sup>23</sup> It is worth noting here that Russell may not have agreed with this assumption for "Dan Brown" as we shall see later. However, I have made this assumption just for the sake of clarity in explaining Russell's theory.

<sup>24</sup> This formulation has been adapted from Salmon/Soames, 1988, p.3.

<sup>25</sup> With the given quote, Russell was referring to the sentence "the King of France is bald".

<sup>26</sup> Frege's answer to this puzzle would probably have been that the apparent meaning of (3) comes from the sense it is expressing, even though it is not denoting a reference.

<sup>27</sup> This similarity is also noted by McKinsey, 1991, pp.154-155; where, roughly, he uses Russell's notion of acquaintance to produce a narrow content which conceptually implies the existence of certain objects "within the agent's immediate mental experience", but does not conceptually imply the existence of certain extramental physical objects. Of course, my reason for bringing in Russell's notion of acquaintance here is different to McKinsey's reason, as I am trying to use it to give the narrow representational content a more "direct" connection or relationship to its external world referents.

<sup>28</sup> In relation to this, I think that the talk of "awareness" in [a] is not meant to imply a *conscious* awareness of a sense-datum, but just that any perceptual experience will involve an interaction with a sense-datum.

<sup>29</sup> This is basically the answer I would give to McCulloch (1995), who, while considering the actions of Liz1 and Liz2, involving water and twin-water, asks "...why should these similarities [the stereotypes of H2O and XYZ] be deemed to be more important than the above differences, unless it is simply being assumed that psychology should be insensitive to differences in real essence?" (1995, p.214). He also declares that "...an Earthian would have to set about things in quite different ways in order to obey the two instructions 'Go and swim in some water' and 'Go and swim in some twin-water'. Obeying the first would still leave the second to do..." (1995, p.214). With this latter point, McCulloch is of course blatantly begging the question against a shared narrow content for Liz1 and Liz2. This seems to be his aim, as from the above, he seems to think that an internalist is begging the question against externalism in relation to the actions of Liz1 and Liz2. However, as I have said, it seems to me that it is the sameness of the duplicates' seemings that is most important in giving adequate explanations for their actions, which will intuitively result in their actions being typed as identical. This to me is not begging the question, but is just being faithful to the subjective viewpoints of the duplicates concerned, i.e., by not bringing in different objective facts which they are unaware of, to explain their behaviour and actions in Twin Earth-type situations.

Indeed, it is interesting to note that the reason McCulloch believes he is not begging the question against the narrow content theorist in the above Liz1/Liz2 example, is because the externalism he supports has been "...established on independent, phenomenological grounds" (1995, p.216). The

externalist phenomenology which McCulloch champions can be described as a Wittgensteinian/Existentialist one, where it seems to a human subject that he or she is embodied and in-the-world, and is interacting with things in the world (1995, pp.131-158). It is this phenomenology that McCulloch believes a Cartesian “film show” or Lockean “snapshot” of experience is incapable of providing. Now, McCulloch might be right about this, but it is important to note that the postulated subjective virtual reality that comes from my narrow representational content, is meant to have the result that it seems to subject concerned that they are embodied and in-the-world. It does seem to the subject that he or she has arms, legs, etc., and is surrounded by his or her environment. This after all, is what the idea of a virtual reality is supposed to encapsulate, at least from the first-person perspective of the subject experiencing it. Moreover, the subject concerned could also appear to be interacting with their virtual environment, perhaps doing a bit of Heideggerian hammering and DIY. Therefore, it could still be claimed that McCulloch is begging the question against narrow content in the Liz1/Liz2 example, as the externalist phenomenology put forward by him could be accommodated by the subjective virtual realities shared by Liz1 and Liz2.

However, even if the foregoing is not accepted, I still don’t think that McCulloch’s externalist phenomenology means that we should necessarily type the actions of Liz1 and Liz2 as distinct. After all, isn’t it plausible that it seems to Liz1/Liz2 that they are embodied and in-the-world, and also interacting with water qua transparent, odourless, liquid etc., rather than water qua H<sub>2</sub>O or XYZ? It seems to be a further semantic step to individuate the waters as different due to their different microstructures.

<sup>30</sup> Even though the narrow content that I have postulated in this chapter originates from visual perception and deals with natural kinds and the physical environment, it seems to me that it could also deal with Burge-type thought experiments involving the social environment of the subject. After all, it is a narrow representational content, which should deal with issues of how the subject represents, or conceives of, things in the world. What I mean by this is that the representational component does not have to consist exclusively of the subject’s sense perceptions; as we have seen, the conceptual role and intensional components can also come into play with this representational content. Briefly, let me illustrate what I mean in relation to Burge’s “arthritis” thought experiment. Actual-Bert and counterfactual-Bert both believe that they have arthritis in their ankles, with the usual difference being that the actual concept of arthritis is a rheumatoid disease of the joints and the counterfactual concept of arthritis is a rheumatoid disease of the joints and bones. Therefore, what would be the narrow representational content of the Berts’ shared belief that they have arthritis in their ankles? Well, as it happens, because arthritis is a disease which causes pain in the subject, we could include a perceptual element with the representational component, as well as the conceptual and intensional elements. For example, it could be claimed that both Berts are representing, or conceiving of, a situation in which the phenomenally identical feelings of pain in their ankles are being caused by the rheumatoid disease called arthritis in their social community, where the shared arthritis-intension will include in its extension a rheumatoid disease of the joints and a rheumatoid disease of the joints and bones. This seems to me the most plausible narrow representational reading of how the Berts are conceiving of things from their subjective points of view.

## CHAPTER 9

### THE POSSIBILITY OF NECESSARY AND SUFFICIENT NARROW CONDITIONS FOR BELIEF CONTENT

Consider the following:

Situation (A): This is a situation in the actual world, where S is perceiving a garden-gnome on a table, which results in S having the belief that there is a garden-gnome on the table.

From the previous chapter, we can postulate that the narrow representational content that constitutes this belief is determined from the proximal stimulations on the visual system of S, which are then utilised in the construction of the internal neural or cognitive representation of the situation in the external world. S instantaneously and almost unconsciously refers to the internal representations of [garden-gnome], [table], etc., with the mental symbols “garden-gnome”, “table”, etc. It is these internal denotations and representations that determine the narrow content of S’s belief that there is a garden-gnome on the table. The question now is whether this narrow representational content can be shown to be both necessary and sufficient for S’s belief having the content that it does.<sup>1</sup>

## 9.1 Necessary and Sufficient Conditions of Narrow Content

This task is going to be both difficult and arguable. Moreover, the difficulties seem to start right from the beginning. For instance, couldn't someone argue that the formulated narrow content from the above actual situation is determined by the state of affairs of there being a garden-gnome on a table in the external world? That is, the internal proximal stimulations used to formulate the narrow content are initially caused by light rays reflecting off the garden-gnome on the table. Therefore, how can we only need narrow content to determine S's belief-content when it seems quite obvious that the wide content from the external world is required to kick-start the whole narrow process?

This issue has arisen before, and, as before, I think that the point one must make is that for an internalist or narrow content theorist, the external world is not considered to constitute or determine the internal representational states of the subject. The external world can certainly be considered as causally responsible for the production of the proximal stimulations on the subject's visual system, but the determination of the internal states is considered to be an internal process, which in the above situation, means that the narrow content of the internal states is determined from the proximal stimulations in. To put the point in Putnam's terms, there is no necessary or logical connection between the events in the external world and the psychological states of the subject. In the above situation, there is only considered to be a contingent causal connection between the external world garden-gnome and table and the resultant proximal stimulations on the sensory receptors of S. Perhaps this point can be illustrated by, for example, the possibility of someone being attached to the earlier

mentioned alien supercomputer, where the proximal stimulations of that person are generated artificially, but which nevertheless results in a neural representation of a state of affairs which is not present in the actual environment of that person.

Moreover, it seems to me that a similar move to the above will be needed for determining whether the narrow representational content of S can be necessary and sufficient for S's belief to have the content that it does. That is, we will need examples where there is a discrepancy or mismatch between the proximal stimulations and internal representations of S on the one hand, and his/her immediate external environment on the other. For instance, consider the following:

Situation (B): This is a counterfactual situation which contains S\*, who is an internal physical/functional duplicate of S from situation (A). S\* is standing in front of a table which has a blue vase on it.

What we have here seems very much like a standard Twin Earth-type thought experiment, with S and S\*'s internal physical states being kept the same and an environmental difference between the actual and counterfactual situations. However, there is a very important difference between the above situations and standard Twin Earth-type thought experiments. The environmental difference between the actual and counterfactual situations is usually one which is at a level below unaided human perceptibility, so that the situations are perceptually identical. But with situation (B), the environmental difference would appear to be perceptible, which in turn seems to cast doubt on whether S\*'s internal states could be identical with S, particularly if S\* is perceiving the blue vase on the table.

However, it will be noted that with situation (B), it is not stated that S\* is *perceiving* the blue vase on the table, he/she is merely standing in front of the blue vase on the table, in a position where he/she ordinarily would be able to perceive it, if all things were equal. But all things are not equal, and this is where the fiendish twist in situation (B) is now revealed. For when I said that S\* is an internal physical duplicate of S, the internal physical states that I had in mind *included the proximal stimulations on the sensory receptors of S*. That is, S\* will have identical proximal stimulations to those of S, where the latter's proximal stimulations were contingently caused as he/she was visually perceiving the garden-gnome on the table in situation (A). I realise that this a highly implausible state of affairs, but there again, examples involving doppelgangers and water which is composed of XYZ do not exactly have an air of plausibility about them either. All I can say in my defence is that a scenario like the above is the only one I can think of to try to tease some necessary and sufficient conditions out of the narrow representational content that has been postulated.<sup>2</sup>

Therefore, what is the intuition that we should have in relation to situations (A) and (B)? It seems to me that there is a plausible intuition supporting the view that the narrow representational content of both S and S\* is identical, i.e., they both have the same internal representation of a garden-gnome on a table. This is despite the fact that in situation (B) S\* is faced with a blue vase on a table.<sup>3</sup> What seems to be the deciding factor in what internal representations S and S\* have is that they have identical proximal stimulations on their visual systems, from which their respective neural representations are constructed.<sup>4</sup> This intuition can be pumped by imagining



that S and S\* are both frozen and attached to the alien supercomputer which then utilises their internal physical states and proximal stimulations to construct their subjective virtual realities. If we further imagine that the formulated virtual realities of S and S\* could be projected onto a viewing screen for all to see, what we would see from situations (A) and (B) would be respective internal virtual realities which contained a garden-gnome on a table.

If the above is accepted, then does the postulated narrow content shared between S and S\* impose any conditions on what their respective belief-contents could be?

Well, a necessary condition for a certain state of affairs is a condition which is required to obtain in order for the state of affairs to occur. In the case of a sufficient condition for a certain state of affairs, this is a condition such that, if it obtains, then the state of affairs will occur, but it may not be a necessary condition for the state of affairs to occur. It may also be the case that a condition is both necessary and sufficient for a certain state of affairs to occur.

### 9.1.1 *A Necessary Condition*

With this in mind, it does seem to me that what situations (A) and (B) show is that a necessary condition for S and S\* to have a belief with the content that there is a garden-gnome on the table is that their respective narrow contents are internally of the type that represents a state of affairs where there is a garden-gnome on a table. For example, it is neither a necessary nor a sufficient condition for a certain external world state of affairs to be in front of S and S\* for them to have a belief-content about

that certain state of affairs. If this was the case, then with situation (B), it would be both necessary and sufficient for S\* to have a belief-content that there is a blue vase on the table. However, the intuition is that S\* does not have a belief with this content, because he/she has the internal representation of a garden-gnome on a table, it therefore seems that he/she necessarily has a belief with the content that there is a garden-gnome on the table.

Further support for the above intuition could also come from the disquotational principle of belief ascription. In situation (A), it seems plausible to suppose that given the narrow representational content of S's experience, he/she would sincerely assent to believing that there is a garden-gnome on the table, the upshot of which would be that we would be entitled to ascribe to S a belief with the content that there is a garden-gnome on the table. In situation (B), it also seems plausible to suppose that given the narrow representational content of S\*'s experience, he/she would sincerely assent to believing that there is a garden-gnome on the table, the upshot of which would be that we would again be entitled to ascribe to S\* a belief with the content that there is a garden-gnome on the table. Moreover, from this it also seems plausible that S and S\*'s behavioural dispositions would be identical, e.g., both S and S\* would be disposed to make the utterance "I believe that there is a garden-gnome on the table" to describe their respective beliefs.

With situation (A), even though the external world state of affairs of there being a garden-gnome on a table is in front of S, it seems to me that this should still not be considered a necessary or sufficient condition for S to have the belief-content that there is a garden-gnome on the table. Recall that this state of affairs was claimed only

to have a contingent causal relationship to the production of the proximal stimulations on S's visual system. That is, it is conceivable that those very same proximal stimulations which results in S having an internal representation of a garden-gnome on a table, and which in turn leads to S having the belief-content that there is a garden-gnome on the table, could have been produced in the absence of the external world garden-gnome being on the table. Therefore, if all of this makes any sense and, moreover, is accepted, then a very powerful claim can be made for the postulated narrow representational content; namely, for a subject to have a belief with a certain content, it is necessary for that subject to have a narrow content which internally represents a certain state of affairs.<sup>5</sup>

What is being claimed above can be stated more formally using a slight variation on Fodor's view of what the Representational Theory of Mind (RTM) would involve for one to have a certain belief. Fodor declares that:

According to the canonical formulation of this view: for any organism *O* and for any proposition *P*, there is a relation *R* and a mental representation *MP* such that: *MP* means that (expresses the proposition that) *P*; and *O* believes that *P* iff *O* bears *R* to *MP* (Fodor, 1985, p.16).

Now, in relation to the above, when Fodor is talking about the mental representation *MP*, I think he means the representation to be constructed out of Mentalese, his mental Language of Thought. However, as I stated previously, my idea of an internal mental representation involves the notion of the representation being formed from the phenomenal elements of the subject's visual experience. The proximal stimulations of S/S\* create an internal mental representation of a state of affairs where there is a garden-gnome on a table. Therefore, if *MP* is understood to be a virtual reality

construction of the distal object or state of affairs in question, then *MP* can serve our purposes adequately in the above formulation. Also, it seems that relation *R* would have to be one which consisted of *MP* being apprehended from the subjective first-person perspective of *O*. If this is accepted, then I think that the above formulation can formally describe what I am claiming about the necessary conditions that the narrow representational content imposes on the consequent belief-content that is formed.

### 9.1.2 *A Sufficient Condition*

Could it also be claimed that a subject's narrow representational content constitutes a sufficiency condition for that subject to have a certain belief-content? This is a difficult issue to decide, as there are a number of different factors which have to be considered. For instance, it seems to me that in situation (B), *S\**'s narrow content, which internally represents a state of affairs where there is a garden-gnome on a table, is sufficient for *S\**'s belief to have the content that there is a garden-gnome on the table. The factor that seems important in deciding this issue for me is how things seem from the subject's first-person perspective. From *S\**'s subjective first-person perspective, the way things seem to him/her is that there is a garden-gnome on the table, which naturally leads to *S\** having the belief that there is a garden-gnome on the table. From this subjective point of view, it seems irrelevant whether there are external world referents for the internal representations of a [garden-gnome] and a [table].

However, for externalists such as Putnam (1981), the question of whether there are external world referents for a subject's internal representations is of vital importance in deciding the content that beliefs could be said to have. For Putnam, S\*'s belief would only have the propositional content that there is a garden-gnome on the table if S\*'s internal representations actually had their relevant external world referents. If this was not the case, then Putnam would claim that S\*'s belief only concerned objects "in the image", such as an image-garden-gnome and image-table, etc., and so it would not be correct to ascribe to S\* the belief that there is a garden-gnome on the table. In his 1981, Putnam was of course referring to the beliefs that a brain-in-a-vat (BIV) could have, but it seems plausible to assume that he would think the same of the scenario involving S\*'s narrowly formulated notional world by the alien supercomputer. In my view, Putnam is here taking an objective third-person perspective on what the content of a subject's belief should be, i.e., it should only be considered "proper" or "correct" belief-content if the belief is about actual objects in the external world. The upshot of this is that a subject's internal representations would not be sufficient on their own for the subject's belief to have a certain content.

With the above positions, it could plausibly be claimed that Putnam's objective third-person approach to belief-content presupposes externalism, whilst my subjective first-person approach to belief-content presupposes internalism. With regards to the issues of belief-content and belief ascription, it seems to depend on which viewpoint is adopted by the ascriber. However, it has been a thread running through this thesis, that in my view it is the first-person perspective of the subject having the belief that should be the determining factor in deciding what content the belief has, or what belief should be correctly ascribed. It is how the subject represents or conceives

things as being, that should provide the individuation conditions for belief-contents, not whether the object or state of affairs that the belief is about actually exists in the external world. Moreover, this latter position has been supported throughout this thesis by plausible appeals to the type-identity of the duplicates' behaviour. If the duplicates have type-identical internal representations of a certain states of affairs, then this would explain their consequent type-identical behaviour.

## **9.2 A First Possible Objection to the Above Arguments for Necessary and Sufficient Narrow Content**

A relevant objection to the above arguments will now be examined. This comes from Davies (1991b), whose aim is to try to show that perceptual content is not locally supervenient, the result of which will be that the determination of perceptual content is viewed as being externalist in nature, rather than internalist. How this is supposed to come about can be explained using situations (A) and (B) above. If S and S\* are considered to be internal physical duplicates, with my claim that the narrow representational content of S in situation (A) is identical to the narrow representational content of S\* in situation (B), despite the environmental differences between situations (A) and (B), then this is equivalent to the claim that the perceptual contents of S and S\* are locally supervenient. By claiming that certain mental properties of a subject are "locally supervenient", what is meant is that those mental properties are fully determined by the subject's internal physical properties. However, if it could be shown that, despite their internal physical identity, the representational contents of S and S\* are different, due to the differences in the

environments of S and S\*, then this would show that perceptual content is not completely locally supervenient, which would be the externalist position that Davies is trying to defend (Davies, 1991b, pp.21-22).

Davies does acknowledge that he is basically trying to defend the externalist position on perceptual content argued for by Burge (1986, 1988a, 1988b), and his arguments are heavily influenced by Burge's thinking. I suppose that I could have concentrated on Burge's argument as an objection to the above claims of necessary and sufficient narrow conditions for representational content. However, the reason that I decided to concentrate on Davies' version of Burge's argument is that he applies it directly to the arguments of McGinn (1989) on perceptual content, which, as I said earlier, did influence my thinking about, and arguments for, the necessary/sufficient narrow conditions for representational content.

### 9.2.1 *McGinn's "Percy" Thought Experiment*

Therefore, in order to examine Davies' argument, it will help to first give a brief description of McGinn's argument. In the actual scenario, a subject called Percy has internal state S1 which is usually caused by distal square things and internal state S2 which is usually caused by distal round things. A counterfactual scenario is then considered, where, perhaps due to different optical laws, Percy's internal state S1 is now usually caused by distal round things and internal state S2 is usually caused by distal square things. As usual in these thought experiments, actual-Percy and counterfactual-Percy have identical internal neurophysical properties (sharing the

same brain states and proximal stimulations), and their behavioural dispositions are also kept the same (McGinn, 1989, pp.63-64).<sup>6</sup> The question now is what is the perceptual content of Percy's internal state S1 in the counterfactual scenario, i.e., is he having an experience as of a square thing or as of a round thing?

McGinn is of the opinion that in the counterfactual scenario, there is a strong intuition that when Percy is in state S1 he will be having an experience as of a square thing, that is, the perceptual content of S1 in the counterfactual scenario will be as of a square thing. This is despite the fact that in the counterfactual world the usual cause of internal state S1 are distal round things. For McGinn, the perceptual content of a state is a phenomenal conception, which involves the notion of how the world seems to the subject (McGinn, 1989, p.58). With counterfactual-Percy in state S1, the intuition is that it seems to him that the world contains a square thing. Moreover, McGinn pumps this intuition in several ways. Firstly, he exploits the fact that the behavioural dispositions of the Percys were kept the same, by claiming that in the counterfactual scenario, when Percy is in state S1, he will be disposed to move square-wise around the distal object, to make square shapes in the air with his finger, etc. (McGinn, 1989, p.64). It is worth noting here that McGinn presupposes that Percy's consequent counterfactual behaviour or actions are also kept the same as in the actual scenario, as it is this presupposition that Davies rejects, as we shall see later. Secondly, McGinn brings in a teleological element, when he claims that "We naturally want to say that the *purpose* of [Percy] moving in a square path is to negotiate square objects successfully, that this is the *function* of his moving like that" (McGinn, 1989, p.66). So it is that McGinn can claim that "A teleological theory of



content may thus be invoked to back up the intuitive verdict I have been urging” (McGinn, 1989, p.67).

In making the above claim for the perceptual content of counterfactual-Percy’s internal state S1, McGinn’s aim is to reject what he calls “strong externalism” about perceptual content. He defines strong externalism as:

...the thesis that a given mental state requires the existence in the environment of the subject of some item belonging to the nonmental world, and that its identity turns on that item (McGinn, 1989, p.7).

In relation to perceptual content, McGinn claims that the strong externalist is committed to claiming that in the above counterfactual scenario, the perceptual content of Percy’s internal state S1 will have changed from what it is in the actual scenario. The reason for this is that as it is a distal round thing in counterfactual-Percy’s immediate environment which is causally responsible for producing state S1 in Percy, then the perceptual content of S1 should be as of a round thing. However, given what has been said above, this claim seems strongly counterintuitive, and I think that most people would agree with McGinn that the intuition is that the perceptual content of counterfactual-Percy’s state S1 is as of a square thing, the same as it was in the actual situation. Indeed, Davies himself also agrees that McGinn has succeeded in rejecting strong externalism for perceptual content (Davies, 1991b, p.36). However, he then goes on to say that “...McGinn’s argument does not at all establish individualism; it does not show that perceptual content is locally supervenient” (Davies, 1991b, p.36).

### 9.2.2 *Davies' Variation of McGinn's "Percy" Thought Experiment*

So it is that we reach Davies' variation of McGinn's "Percy" thought experiment, which he claims shows that the perceptual content of counterfactual-Percy's internal state S1 is different to what it was in the actual situation, the upshot of which is that it appears that perceptual content is not locally supervenient. It is also this variation that may provide an objection to my above claims for necessary and sufficient narrow conditions for representational content. Davies describes the variation thus:

In the counterfactual situation we now find, not Percy himself, but a duplicate with a very different evolutionary history. This creature's ancestors survived to reproduce in part because their behaviour was appropriate to the distal causes of their perceptual experiences. In this imaginary scenario, internal state S1 is produced by distal round things, as in McGinn's example; but the behaviour consequent upon the creature's being in S1 is now appropriate to the presence of round things, and not to the presence of square things...we suppose that environmental differences have the consequence that the same nerve firings and muscle contractions as in the actual situation result in a quite different bodily trajectory. In particular, the goings-on inside the skin which in the actual situation lead to a square trajectory now have a round trajectory as their upshot. This happy agreement of input-side, output-side, and teleological factors makes it plausible that, when Percy's duplicate is in state S1, he has an experience as of a round thing. A fortiori, it is implausible that the duplicate misperceives round things as square (Davies, 1991b, pp.37-38).

From the above quotation, it seems to me that the crux of Davies' claim is the manner in which the behaviour of Percy's duplicate is now considered appropriate to the distal object causing internal state S1. With McGinn's thought experiment, the fact that the consequent behaviour of counterfactual-Percy being in state S1 was considered inappropriate to his environment was a major factor in pumping the intuition that the perceptual content of state S1 was as of a square thing. Moreover, this intuition was further supported by teleological factors, i.e., it was a function of counterfactual-Percy's square-wise behaviour to avoid square objects. However, with

Davies' variation, the appropriate round-wise behaviour is now pumping the intuition that the perceptual content of S1 is as of a round thing, and this intuition is now being supported by teleological factors.

In changing the consequent behaviour of Percy's duplicate being in state S1, has Davies made a legitimate move? After all, Percy's duplicate is still presumably a neurophysical/functional duplicate, and more importantly, the behavioural dispositions of Percy's duplicate are also kept the same as actual-Percy. Therefore, if this is the case, how can the behaviour of Percy's duplicate differ to that of actual-Percy? The answer to this was alluded to earlier, and seems to depend on the difference between behavioural dispositions and overt behaviour. As Davies explains:

The externalist is allowed to have the duplicate's behavioural dispositions differ from those of the actual subject, to the extent that this is consistent with the two being duplicates...This may appear to be a negligible degree of freedom for the externalist since, surely, the basis of behavioural dispositions is to be found inside the skin. But, if behaviour is itself characterised externalistically, then the production of behaviour of a certain type depends both upon what happens inside the skin – nerve firings, muscle contractions, and the like – and upon environmental factors. In principle, behaviour – externalistically characterised – can be varied even while everything inside the skin remains the same (Davies, 1991b, pp.32-33).

Therefore, in his variation of McGinn's example, what Davies is claiming is that when in state S1, Percy and his duplicate are still duplicates with regards to the internal behavioural dispositions they have, i.e., they both have the disposition to square-wise behaviour. However, due to, say, environmental differences in the counterfactual situation, the overt or externalistically characterised behaviour of the duplicate is considered different to that of actual-Percy, and is appropriate to the

counterfactual environment, i.e., it is round-wise behaviour. Perhaps the externalist could claim that this move is equivalent to McGinn's original ploy of having a distal round thing counterfactually producing the same internal state S1 which is produced by distal square things in the actual situation?

For the sake of the argument, let us agree that Davies' strategy of externalistically characterising the behaviour of Percy's duplicate is legitimate. The question that must now be addressed is whether this move produces a plausible thought experiment that shows that the perceptual content of the counterfactual state S1 is different to that of the actual state S1, which would then show that perceptual content is not locally supervenient.

Personally, I don't think that Davies has produced a plausible thought experiment to this end; indeed, the more I think about it, it seems a totally implausible thought experiment. What leads me to this claim of implausibility is if we consider the subjective first-person perspective of Percy's counterfactual duplicate. To begin with, the duplicate shares an internal physical/functional identity with Percy, which includes the proximal stimulations on the sensory receptors. Therefore, we can say straight off that from the subjective viewpoint of the duplicate, when he is in counterfactual state S1, he is having a visual experience as of a square thing. I believe I can plausibly claim this because the internal representation produced from the proximal stimulations will produce a square shape which appears to the duplicate to be located and extended in space, the same as it does for Percy. Moreover, because the duplicate shares a functional identity with Percy, this means that he shares a conceptual identity with Percy, which will mean that he possesses the concept of

square, and so does recognise that the internal representation he is experiencing while in state S1 is as of a square thing.

Now, I don't think that Davies would disagree with the above, the reason being that he seems to be counting on an attribution of perceptual content to Percy's duplicate that comes from the objective third-person perspective. That is, mainly due to taking account of the externalistically characterised behaviour of the duplicate while in S1, which is appropriate to his environment, i.e. appropriate to there being a round thing in his immediate environment. This is then backed up by a teleological appeal to a successful evolutionary history for his ancestors and their behaviour while in S1. Thus it is that Davies believes we should attribute the perceptual content "as of a round thing" to the duplicate when he is in S1.

However, it seems to me that to accept such an attribution of perceptual content would involve completely ignoring the first-person perspective of Percy's duplicate; it is as if one is throwing the subject out with the Cartesian bath-water. Indeed, Davies seems to downplay what is involved with the subjective viewpoint of Percy's duplicate. For example, in talking of the behavioural dispositions that presumably Percy and his duplicate share, he describes them as "nerve firings and muscle contractions" (Davies, 1991b, p.34), which gives the impression that there is no subject involved in this at all, just objective biological events and processes.

However, if the behavioural dispositions of Percy's duplicate are described from the subjective viewpoint, a different picture emerges. For instance, consider the duplicate being in state S1. Even though there is a distal round thing in front of him, from his subjective viewpoint he is having a visual experience as of a square thing, as I

claimed above. But now consider his behavioural disposition to, say, trace a square shape in the air using his finger. From the subjective viewpoint of the duplicate, he does trace a square shape in the air. Moreover, if we ask him what shape he has traced, from his subjective viewpoint, he does make the utterance “A square”. But of course, all this counts for nothing from the objective viewpoint of an observer, who, due to the behaviour being externalistically characterised, sees the duplicate trace a round shape in the air with his finger and utter “A circle”.

This is what I mean when I claimed that the subjective first-person perspective of Percy’s duplicate was being ignored. Indeed, with the set-up that Davies is suggesting for attributing perceptual content, you don’t even need the subject to be there in the first place. Instead of a duplicate of Percy in the counterfactual scenario, one could have a robot duplicate of Percy that had no inner subjective life at all. This wouldn’t affect the attribution of perceptual content from the objective perspective at all. The internal state S1 of the robot-Percy would now be a computational state, but could still be attributed the perceptual content of “as of a round thing”, due to the robot’s overt behaviour of tracing a round shape in the air with it’s finger and uttering the words “A circle” on cue.

What all this seems to show is that there is a behaviourist or instrumentalist aspect to the method that Davies is proposing for attributing perceptual content to Percy’s counterfactual duplicate. Of course, McGinn also used behavioural dispositions and overt behaviour in his original Percy thought experiment. But with McGinn’s thought experiment, the internal representation of a square thing was already there, constructed from the proximal stimulations of counterfactual-Percy. McGinn was

simply using the square-wise behaviour of counterfactual-Percy to pump the intuition that the perceptual content of his internal state S1 really was as of a square thing. With Davies' variation, the overt behaviour is being used to override the perceptual content that state S1 seems to have from the subjective viewpoint of Percy's duplicate.

Therefore, in Davies' variation, it seems to me that if the first-person subjective perspective of Percy's duplicate is given due attention while the duplicate is in state S1, it seems a far more plausible prospect that, contra Davies, the duplicate is misperceiving round things as square. The perceptual content of the counterfactual state S1 is the same as the perceptual content of actual state S1 – as of a square thing, and that seems to show, contra Davies, that perceptual content is locally supervenient.<sup>7</sup>

### *9.2.3 How Davies' Variation of McGinn's Thought Experiment Could be Construed as an Objection to My Arguments*

Having examined Davies' variation of McGinn's "Percy" thought experiment, it seems clear enough how the strategy involved with this variation could be used as an objection against my above arguments for necessary and sufficient narrow conditions for representational content. Recall situation (B) where S\* is standing in front of a table with a blue vase on it. I argued that because S\* had the same proximal stimulations on his visual receptors as S had in situation (A), the former would not have an internal representation of a blue vase on a table, even though that was the

distal state of affairs in front of him. It was argued that S\* would have an internal representation of a garden-gnome on a table (the same as S), which was the basis for his consequent belief that there is a garden-gnome on the table.

However, one can imagine a Davies-type variation of situation (B), where, for example, it is claimed that the distal occurrence of a blue vase has always produced an internal representation of a garden-gnome in S\*. This can be followed up by externalistically characterising the behaviour of S\*, so that it is in harmony with his immediate environment. So, for example, when asked to describe his belief in situation (B), due to the effects of gravity or friction, etc., S\* would make the utterance “I believe that there is a blue vase on the table”. In this manner, the intuition would be pumped that the representational content of S\* in situation (B) is different to that of S in situation (A), despite their internal physical/functional identity, i.e., S\* would have a representational content as of a blue vase on a table.

However, as with Davies’ variation of McGinn’s “Percy” thought experiment, it seems to me that a consideration of the first-person subjective viewpoint of S\* in the variation of situation (B) would motivate the intuition that S\* does indeed have an internal representation with the content as of a garden-gnome on a table. For instance, the behavioural dispositions of S and S\* are held constant, so from the subjective viewpoint of S\* he intends to utter, and does believe that he has uttered, the sentence “I believe that there is a garden-gnome on the table” when asked to describe his belief, even though this is externalistically “squashed” so that it comes out as him believing that there is a blue vase on the table. Once again, the first-person perspective of the subject is being completely ignored with the third-person



attribution of representational content. Therefore, it seems to me that a Davies-type variation of situation (B) would be just as implausible for fixing perceptual or representational content as Davies' variation of McGinn's "Percy" thought experiment was.

#### 9.2.4 McGinn's "*Weak Externalism*" and *Narrow Representational Content*

It is interesting to note that the upshot of McGinn's "Percy" thought experiment could be described as being a situation where actual-Percy and counterfactual-Percy shared an internal state that had the same narrow representational content. After all, narrow content has been defined in this thesis as a mental content that is shared between doppelgangers in Twin Earth-type situations. However, even though McGinn rejects strong externalism for perceptual content, he still takes the view that perceptual content is weakly externalist in nature. He defines "weak externalism" as:

...the thesis that a given mental state requires the *existence* of some item belonging to the nonmental world, and that its identity turns on that item (McGinn, 1989, p.7).

As we have seen, in relation to perceptual content, strong externalism requires that for a subject to have a mental state about a particular object, the object must be in the immediate external environment of the subject and there must typically be some sort of causal relation between the subject and the object. With weak externalism about perceptual content, this requires that for a subject to have a mental state about a particular object, the object must exist somewhere in the external world, it does not have to be in the immediate environment of the subject and there need be no causal

interaction between subject and object.<sup>8</sup> What I think this means is the following. In the counterfactual situation of the “Percy” thought experiment, where Percy’s internal state S1 has the perceptual content as of a square thing even though it is caused by distal round things, that content is only possible, and is only weakly external, because there are distal square things existing somewhere in the external world of Percy. And, of course, as McGinn has described his thought experiment (McGinn, 1989, pp.63-64), there are distal square things in the counterfactual external world of Percy (only these are now producing internal state S2 in Percy, which has the perceptual content as of a round thing).

However, in relation to the perceptual content involved with counterfactual-Percy, McGinn pushes the weak externalist line slightly differently, talking of “...(weak) externalist descriptions of *behaviour*: we need to conceive of behaviour in terms of properties that objects might...instantiate” (McGinn, 1989, p.69). So the perceptual content of S1 – as of a square thing – is considered weakly externalist because Percy’s consequent square-wise behaviour involves the property of being square, which is not an internal property of Percy (McGinn, 1989, p.68). It is these weak externalist descriptions of behaviour that McGinn claims enable him to escape a tricky situation, that is:

...how can weak externalism be true of perceptual content if such content is supervenient upon ‘internal’ facts of the three kinds I have distinguished? [That is – brain states, proximal stimulations and behavioural dispositions] That is, if it is sufficient for a certain kind of perceptual content that these internal conditions obtain, then how can it also be *necessary* to include some objective property in the specification of that content? Surely, if perceptual content is so supervenient, then we must be complete internalists about perceptual content, thus denying objective content to perceptual representations (McGinn, 1989, p.69).

What McGinn claims here is, I think, equivalent to what I said above, concerning the point that actual-Percy and counterfactual-Percy seem to have identical narrow representational contents. Indeed, the necessary and sufficient narrow conditions for representational content that I have argued for, are I think, the position that McGinn calls “a disaster” (McGinn, 1989, p.69). But does McGinn’s weak externalist strategy, involving both existing external objects and externally described behaviour, enable him to escape the “disaster”? I think it does, given the way that McGinn has defined weak externalism for perceptual content, i.e., counterfactual-Percy’s consequent behaviour does entail the need for there to be the property of squareness in the external world.

However, we don’t have to go along with McGinn’s definition of weak externalism for perceptual content. The narrow representational content which I have been postulating in this part of the thesis, does seem to be determined only by the “internal” neurophysical facts of the subject and his behavioural dispositions. Moreover, I have suggested that this representational content which is formed from the proximal stimulations in, appears from the first-person perspective of the subject as a narrow virtual reality of the external world. The upshot of this for the subject, is that his behavioural dispositions do appear to translate to his consequent behaviour, but this apparent behaviour is not in the external world, but takes place in the virtual reality of his narrow representational world.

It is by this method, that the subject’s behavioural dispositions don’t just have to be defined as nerve firings and muscle contractions, i.e., non-intentionally, but can have the quasi-intentionality that comes from his narrow virtual environment. For

example, with counterfactual-Percy and his square-wise behaviour, this need not entail the existence of the property of squareness and square objects in his external world, it just entails the existence of the property of squareness and square objects in his narrow virtual reality of the external world. So even though the property of squareness is not an internal property of Percy, qua physical Percy, it is a property of his internal neurophysically constructed representational world. Therefore, if this is accepted, for Percy to have a perceptual content that results in square-wise behaviour, this does not entail that the property of squareness or square things have an objective existence in the external world of Percy.<sup>9</sup> So, it seems a possibility that one person's "disaster" can be another person's narrow content. Therefore, following McGinn's definitions, perhaps a definition can be given of narrow representational content as:

Mental states which require the existence of some item belonging to the internally constructed (mental) virtual world of the subject, and that their identity turns on that item.

### **9.3 A Second Possible Objection to the Above Arguments for Necessary and Sufficient Narrow Content**

There is another objection that could be made to the above arguments for a necessary and sufficient narrow content. Indeed, this objection could also have been raised in the previous chapter, as I think it is also applicable to situation (i), where I had the actual and counterfactual internal physical duplicates perceiving a glass of water on a table. The aim of this was of course to try to work out what narrow representational

content was shared between them during these perceptual episodes, even though the respective waters had different chemical compositions, i.e., H<sub>2</sub>O and XYZ. However, I think that here is the appropriate place to deal with this objection, and so I will examine its effect on both the above argument for necessary and sufficient narrow content and the previous situation (i).

The objection itself comes from the philosophical position called wide externalist representationalism.<sup>10</sup> The notion of representationalism is one which concerns the qualitative or phenomenal character of conscious mental states, where this involves “what it is like” for the subject as he/she experiences the conscious mental state. For example, the phenomenal character of a subject’s visual experience of a red balloon would be that it is an experience that has a particular reddish character to the subject having it. Moreover, on a representationalist construal of phenomenal character, it would be said that the phenomenal character of the visual experience of the red balloon would involve representing the intentional object of the experience as having a certain property, i.e., it represents the balloon as having the property of being red (Levine, 2003, p.58).

There are a number of different types of representationalism, which are well explicated by Levine (2003) (where instead of representationalism, he refers to “intentionalism”, but they mean the same thing):

Wide intentionalism is the doctrine that qualitative character is to be identified with wide intentional content, whatever that turns out to be. Obviously, then, narrow intentionalism is the doctrine that it is narrow content with which we identify qualitative character...Most adherents of wide intentionalism are also externalists, in the sense that the wide contents they attribute to qualitative

experiences are the relevant physical properties of the distal objects of perception (Levine, 2003, p.59).

Therefore, given the above, I think that we can say the following. On a wide representationalist reading of the visual experience of the red balloon, the qualitative character of the experience would involve representing the distal balloon as having a reddish property. In contrast, on a narrow representationalist reading of the visual experience of the red balloon, the qualitative character of the experience would (I think) involve representing the proximal balloon of the actual experience as having a reddish property. On the former reading, the qualia of the experience would be determined by the external world balloon, whilst on the latter reading, the qualia of the experience would be determined by the internal representation of the balloon.

If we now continue following Levine's distinctions, we can say that on a wide externalist representationalist reading of the visual experience of the red balloon, the qualia of the experience would not just be determined by the distal balloon, but would be identified with the physical properties of the balloon, e.g., the surface reflectance properties of the balloon. Levine doesn't mention what a narrow internalist representationalist reading would involve, but perhaps we can say the following in relation to the visual experience of the red balloon. It would involve the qualia of the experience being determined by the internal representation of the balloon, but they would be identified with the internal physical properties of the subject, where this would involve the proximal stimulations of the visual receptors and the neural states of the subject's brain. I'm not sure about this, but it does seem quite amenable to the position I have been arguing for previously, so I think it's safe to say that the position I most identify with here would be that of narrow internalist representationalism.

### 9.3.1 *The Possibility of Absent and Inverted Qualia*

Nevertheless, one might ask what all this has to do with the arguments for necessary and sufficient narrow conditions for belief content given above and the duplicates in situation (i) in the previous chapter. However, it is the consequences that follow from wide externalist representationalism that impact on the positions that have been argued for in relation to my construal of narrow representational content. As Gray (2003) puts it:

Externalist forms of representationalism...reject the claim that the phenomenal character of experience...supervenes on the neural. If the phenomenal character of experience does not supervene on the neural then the question arises of whether neural duplicates might differ phenomenally (e.g. whether their 'qualia' might be inverted). Further, and even more contentiously, the question arises of whether a neural duplicate of a sentient creature might lack experience altogether (i.e. whether its 'qualia' might be absent) (Gray, 2003, p.245).

Therefore, the consequences that come from wide externalist representationalism are the possibilities of inverted qualia and absent qualia. This sounds bizarre, but it does follow logically from what has been described above. If wide externalist representationalism identifies the qualitative character of a sensory experience with the physical properties of the distal object that is being perceived (or the intentional content), then any changes in that distal object (or intentional content) should result in changes to the qualia of the sensory experience.

However, it seems to me that these consequences are extremely counterintuitive and implausible. Firstly, consider the case of absent qualia. If wide externalist representationalism is correct it would mean that if two neural duplicates were both

having a visual experience of a red balloon, but that a distal red balloon was only present in front of one of the duplicates, then the result of this would be that the duplicate who did not have a distal red balloon in front of him/her would not be having an experience with any qualitative character. However, this is deeply implausible, especially if we construe the “qualitative character” of the experience as how the experience seems to the subject having it. Even though there may not be a distal red balloon in the environment of one of the duplicates, there is an intuitive plausibility (due to the theses of local mind-brain supervenience and first-person authority) that the duplicate will still be having a visual experience of a red balloon, and that this experience will have a qualitative character, i.e., it will seem to the subject that he/she is having an experience with a reddish character.<sup>11</sup>

We now come to the case of inverted qualia. This possibility comes about because if, according to wide externalist representationalism, the qualitative character of a sensory experience is determined by the distal intentional content of the experience, this means that if that distal intentional content changes, there should also be a change in the qualia of the sensory experience. This sounds reasonable enough, but a problem is raised by Block’s (1990) “Inverted Earth” thought experiment. Inverted Earth is a place where the colours of things are inverted to the complementary colours of the colours on Earth (Block, 1990, p.489). For example, a red balloon on the Earth, would be a green balloon on Inverted Earth. Block now imagines that some mad scientists knock a person out, and while unconscious, they insert inverting lenses in that person’s eyes. This person is then taken to Inverted Earth and wakes up there, and due to the inverting lenses in his/her eyes, everything appears as it did on the Earth. So when he/she has a visual experience of what is really a green balloon, it



seems to the person that the qualitative character of the experience is representing the balloon as having a red property.

However, it is intuitive to think that after some time has passed, the intentional contents of the person's conscious sensory experiences will change. For example, whereas on the Earth, the person's reddish visual experiences generally have intentional contents that are red, on Inverted Earth, the person's reddish visual experiences generally have intentional contents that are green. So we have a situation where the qualitative character of the person's visual experiences has stayed the same, but the distal intentional contents of those experiences have changed. For Block, "[t]his is enough to refute the functionalist theory of qualitative content and at the same time to establish the intentional/qualitative distinction" (Block, 1990, p.490). The latter distinction that Block mentions, is of course due to the fact that if wide externalist representationalism is correct, because the distal intentional contents of the person's reddish visual experiences have changed on Inverted Earth, then the qualitative character of those experiences should be inverted, i.e., the person should now be having greenish visual experiences. However, this seems implausible, as the person concerned notices no change to his/her reddish visual experiences.

This result of Block's Inverted Earth thought experiment has been disputed, and it may be instructive to examine some of the arguments put forward in support of externalist representationalism. One of the latter's staunchest defenders has been Michael Tye (1995, 1998, 2000). For example, Tye (1995) appeals to evolutionary factors to claim that the intentional or representational contents of the person's visual experiences would not change on Inverted Earth, and so his/her qualia would not be

inverted. Tye claims that "...on Inverted Earth, optimal conditions do not obtain. The brain state in me that tracks blueness in optimal conditions (and thereby represents blueness) now tracks yellowness. But it does not now *represent* yellowness (Tye, 1995, p.207). It is only in the natural habitat of Earth that optimal conditions are present for the person's brain state to correctly track and represent the colours of his/her visual experiences. However, isn't it plausible to think that after, say, 10 years, Inverted Earth would have become the inverted-lens person's natural habitat, with his/her brain state correctly tracking blueness which now represents yellowness, which would mean, according to externalist representationalism, that the blue qualia of the person would invert to yellow qualia? Perhaps Tye could respond to this by saying that because the person had inverting lenses attached, there never would be a natural habitat, and so, optimal conditions, for a change of intentional content to occur (Tye, 1995, p.207).

(Interestingly, this latter response is again put forward in Tye (2000) as the reason that the representational content of a traveller's inverted lens visual experiences of the Inverted Earth sky would not change. The upshot of this is that even after years have passed on Inverted Earth, the traveller's visual experience of the sky would still have blue qualia as the representational content of the experiences is still the property of blue. However, Macpherson (2005) raises a problem for Tye's position. The problem comes about because even though Tye claims that the qualia and content of the traveller's visual experiences of the Inverted Earth sky have not changed, he does claim that the belief-contents that the traveller has will have changed for the standard externalist reasons put forward by Putnam and Burge. This means that when the traveller has the belief that he/she expresses with the sentence "The sky is blue", what

he/she now means by this is “The sky is yellow” (Macpherson, 2005, p.138).

Therefore, because the traveller’s visual experience of the sky still has blue qualia and blue content, the bizarre outcome of this is that, as Macpherson puts it, “The traveller seems to be radically in error about the nature of their current experience”

(Macpherson, 2005, p.138)).

Nevertheless, Tye (1998) acknowledges that his earlier evolutionary response to Inverted Earth faces a problem. It would entail that the “Swampman” from Davidson (1987) would have no conscious experiences (Tye, 1998, p.461). Swampman is considered to be an internal physical duplicate of an existing human being, who is created instantaneously by a lightning strike in a swamp (perhaps he looks a bit like George Bush, only more hairy). Because Swampman has no evolutionary history, Tye claims that “His inner states play no teleological role. Nature did not design any of them to do anything. So, if phenomenal character is a certain sort of teleo-representational content, then Swampman has no experiences” (Tye, 1998, p.461). Tye finds this conclusion counterintuitive, and thinks it more plausible to imagine that Swampman would have phenomenal experiences, perhaps due to his brain states correctly tracking representational content “...under conditions of *well-functioning*” (Tye, 1998, p.463).

However, this now leads to a dilemma for externalist representationalism, for if it is allowed that Swampman can have phenomenal sensory experiences by a non-teleological method, this then opens up the possibility of Swampman having inverting lenses attached to his eyes and being unknowingly transported to Inverted Earth. The upshot of this is that Swampman will notice no difference to the phenomenal

character of his visual experiences of, say, the blue sky, even though the objective colour of the sky on Inverted Earth is actually yellow. However, the externalist representationalist will not now be able to appeal to teleological evolutionary factors to claim that the representational content of Swampman's visual experiences of the sky has not changed when he is transported to Inverted Earth. This in turn means that according to externalist representationalism, the representational content of Swampman's visual experiences of the sky will eventually change, with the result that the visual qualia of his experiences should also change, from blue qualia to yellow qualia. However, this seems highly implausible as due to the inverting lenses, Swampman notices no change in the phenomenal character of his visual experiences of the sky, i.e., they remain blueish visual experiences (see Tye, 1998, p.464).

Tye's response to this dilemma is an interesting one. He tries arguing that Block is actually begging the question by claiming that the Earth person with inverting lenses would have identical phenomenal experiences when on Inverted Earth. He does this by claiming that this situation would involve the person's memories of phenomenal experiences, which may well be externalist in nature, in the same way that a person's present water-beliefs may be different to his/her water-beliefs in the past, without the person realising this. A situation like this would occur if the person concerned had originally been on Earth where water is H<sub>2</sub>O, but is now on Twin Earth where water is XYZ (Tye, 1998, pp.464-472).

I must admit that Tye argues his case well, and it does make one think that Block may well be begging the question on the issue of a person having identical phenomenal experiences on Earth and Inverted Earth. However, it must be remembered here what

the bottom-line is for Tye, what the conclusion is that he is arguing for. Even if one agrees that Block is begging the question, Tye is trying to argue that for a Swampman with inverting lenses on Inverted Earth, the phenomenal character of his visual experiences of the sky will change due to the change of representational content of those visual experiences, but that Swampman will not notice this! So Swampman starts off by having blue visual experiences of the Inverted Earth sky, but after some time has passed, he then has yellow visual experiences of the Inverted Earth sky, but does not notice this change of visual qualia. It is when Tye's conclusion is baldly stated like this, that unfortunately for him, a tidal wave of implausibility comes crashing over his position. Tye does claim that this change would occur over a period of time and only in switching cases (Tye, 1998, p.471), but I don't think this helps much. Tye is not just biting the bullet here, he's trying to chew on a cruise-missile.

The irony here is that whereas with content externalism, most of the compelling thought experiments are in support of it, with externalist representationalism, most of the compelling thought experiments seem to argue against it. When externalist representationalism ties the phenomenal character of a visual experience to the physical surface of the distal object which is the intentional or representational content of the experience, this leads to claims of necessity which appear too strong to uphold. For instance, Block's Inverted Earth thought experiment is a counterexample to the externalist representationalist claim that if a subject's two visual experiences have different representational contents, then necessarily, those visual experiences will have different phenomenal character. The closely related claim to this is that if a subject's two visual experiences have the same representational contents, then necessarily, those visual experiences will have identical phenomenal character.

However, this latter claim is also open to attack by counterexamples. For instance, Levine (2003) imagines a situation where creatures have eyes on the sides of their heads like fish, so that their visual fields don't overlap, so they cannot look at the same object with both eyes (Levine, 2003, p.69). He further imagines that "...the lenses of the two eyes were colour-inverted with respect to each other" (Levine, 2003, p.69). The upshot of this is that it is plausible that the creature could be having simultaneous visual experiences with identical representational content (i.e., distal objects with identical physical surface reflectances), but which would result in the visual experiences having distinct phenomenal character. This result is of course contra to the claims of externalist representationalism, but it is a result which is far more plausible and intuitive than the claim that the qualia of the creature's visual experiences are really the same, so that the creature is mistaken in its judgement that the qualia are distinct (Levine, 2003, pp.69-70).

### 9.3.2 *How The Consequences of Wide Externalist Representationalism Impact on My Positions in Situation (i) and for Necessary/Sufficient Narrow Content*

#### Situation (i)

Having explicated the wide externalist representationalist position and some of the consequences that flow from that position, it is time to examine what impact these consequences have on the positions I have been arguing for. Firstly, let us go back to situation (i) in the previous chapter, where we have the internal physical duplicates S and S\* both visually perceiving a glass of water on the table. At this point, I think

that everyone, i.e., internalists, externalists, and anyone hovering in between, would be satisfied that S and S\* are having identical conscious visual experiences. On a representationalist reading, S and S\* are having identical visual experiences with a phenomenal character that is representing there being a glass of water on a table. This means that it seems like something to the subjects concerned while they are having their visual experiences, namely, it seems to them that they are having a visual experience which represents there being a glass of water on a table.

In particular, wide externalist representationalists will be satisfied, as in both scenarios, there actually is a distal intentional content to the experiences, which is the external world glass of water on a table. It is the physical properties of this distal intentional content that the wide externalist identifies with the qualia of the visual experiences, so as long as it is there, the externalist can agree that the phenomenal character of the visual experiences are identical for S and S\*.

It is the next move in the situation that causes friction and questions to be asked about the identity of S and S\*'s conscious visual experiences, where it is imagined that S and S\* are frozen mid-visual experience and attached to the alien supercomputer, where the latter then "restarts" the respective experiences artificially, only utilising the neurophysiological properties of the subjects (i.e., proximal stimulations and brain states). Of course, this is far-fetched, but by imagining S and S\* being frozen mid-visual experience and removed from their immediate environments, I was trying to gauge the intuition as to how much of their environment would come away with them when they were frozen and removed from it. A weak analogy would perhaps be a pin-hole camera which is left in front of, say, a tree, and then removed after a while to

find out how much of an image of the tree had been formed inside it. Of course, the pin-hole camera is passive, whereas our neurophysiology is very active in constructing the reality we experience, but it's something like what I had in mind.

Nevertheless, when S and S\* are attached to the alien supercomputer and their glass of water/table experiences are then reactivated artificially, what is the intuition with regards to the identity or otherwise of their conscious experiences? For wide externalist representationalists, this situation is going to result in them claiming that the duplicates are no longer having visual experiences with identical qualia. Indeed, it is not so much a case of inverted qualia, but they will argue that the duplicates are not having any conscious visual experiences at all, a case of absent qualia. This of course follows from the fact that S and S\*'s visual experiences whilst attached to the alien supercomputer do not have their distal intentional contents with accompanying physical surface properties anymore, which is where the qualia of their visual experiences are determined. (Wide representationalists would also probably balk here because the visual experiences do not have their distal intentional contents anymore, which is where they believe the qualia of the experiences are determined).

But how plausible is this position? Are we really to believe that when S and S\* have been attached to the alien supercomputer and their respective visual experiences of a glass of water on a table are reactivated, that they have no qualitative visual experiences at all? To me this seems highly implausible. The intuition I get here is that whilst in situation (i), due to S and S\* being neural duplicates, they are having qualitatively identical visual experiences of a glass of water on a table. On a representationalist reading, we could say that they are having identical visual



experiences which are representing the intentional contents of those experiences as being a certain way, i.e., that there is a glass of water on a table. Now, in the veridical situation (i), are those intentional contents distal ones or proximal ones? Or is it the case, to stay consistent with what I have been arguing for in this and the previous chapter, that it is a case of the proximal intentional content of the experience being superimposed over the distal intentional content of the experience? I suppose that this would be a narrow representationalist position, where the qualia of the visual experiences are determined by the internal representations, through which we then experience the external world reality.

Of course, in the veridical situation this would not be noticed, but when S and S\* are frozen mid-experience and attached to the alien supercomputer, it is only then that the internal part of the visual process becomes apparent. S and S\* are still neural duplicates and if the plausible idea of the local supervenience of mental properties on neurophysiological properties counts for anything here, they will continue to have qualitatively identical visual experiences which represent there being a glass of water on a table. Moreover, due to the virtual reality nature of these experiences, the glass and the table will appear to be located and extended in space, this is not a Lockean image or picture that I am talking about. The experiences will seem as they were in the veridical situation except that from an objective third-person perspective, the external world reality is no longer under the proximal representations of it.

Therefore, in situation (i) and when attached to the alien supercomputer, I find it intuitively plausible that the visual experiences of S and S\* are intentionally identical, in that they are both representing there being a glass of water on a table, and also

qualitatively identical, in that it seems to both S and S\* that they are having visual experiences of a glass of water on a table.<sup>12</sup> So it was legitimate for me to utilise these qualitatively identical visual experiences of S and S\* to try to find out what narrow representational content is shared between them.

### Necessary/Sufficient Narrow Content

It is now time to consider what impact the conclusions of wide externalist representationalism has on my argument for the necessary and sufficient narrow conditions for belief content. This involves situation (A), where S is having a veridical visual experience of a garden-gnome on a table. We then move to situation (B), where S\*, who is an internal physical duplicate of S is standing in front of a table which has a blue vase on it. S\* is in such a position that if all things are equal he/she would be having a veridical visual experience of a blue vase on a table. However, in this situation, it has been suggested that S\*'s proximal stimulations on his/her visual receptors have been fixed to what they were with S in situation (A), when S was visually experiencing the garden-gnome on a table. From this I suggest that the intuition is that in situation (B), S\* is having a visual experience as of a garden-gnome on a table, rather than a visual experience of the blue vase on a table. I then try to argue that this narrow representational content which is shared between S and S\* is necessary and sufficient for S and S\* to have respective beliefs with the content that there is a garden-gnome on a table.

It seems clear here that the objections to my above claims from wide externalist representationalism are going to be similar to the objections that were made against the identity of the conscious visual experiences of the duplicates in situation (i), which have just been examined. In situation (A), wide externalist representationalists will probably be satisfied that the phenomenal character of S's visual experience is representing a situation where there is a garden-gnome on a table. This is because the qualia of the visual experience can be determined by the physical properties of the distal intentional content, e.g., the surface reflectance properties of the garden-gnome on a table.

However, with situation (B), wide externalist representationalists are going to object that the visual experience of S\* does not have an identical phenomenal character to the visual experience of S, even though the proximal stimulations on the visual receptors of S and S\* are identical. This is due to the fact that S\*'s visual experience as of a garden-gnome on a table now lacks the appropriate distal intentional content, which is of course what determines the qualia of visual experiences for externalist representationalists. Indeed, this seems to imply that in situation (B), S\*'s visual experience would have no phenomenal character at all, it would be a case of absent qualia. Therefore, the visual experiences of S and S\* would not have identical phenomenal character and so would not be identical visual experiences.

In answer to this objection, I will state what I understand to be the phenomenal character of a visual experience. It is the "what it is like" factor for a subject having a visual experience, it entails that the experience seems a certain way to the subject who is having it. The phenomenal character of a visual experience is something which can

be apprehended only from the subjective first-person perspective of the subject having it, and is not available to the objective third-person perspective, as has been argued by Thomas Nagel et al. If this is accepted, then for externalist representationalists to argue that in situation (B), S\* is not having a conscious visual experience seems to me completely counterintuitive.

The intuition which has been of vital importance to the notion of narrow content in this thesis has been that what a subject believes is wholly supervenient on his or her internal physical properties, how the subject is on the inside. I believe that this is true even more strongly for the phenomenal character of a subject's sensory experiences, how those experiences seem to the subject having them. The fact that there is not the appropriate distal intentional content for S\*'s visual experience in situation (B) seems to me irrelevant to how the experience seems to S\* from the first person perspective. It is intuitively plausible to me that in situation (B) it seems to S\* that he/she is having a conscious visual experience of there being a garden-gnome on a table, i.e., in representationalist terms, it seems to S\* that he/she is representing a situation where there is a garden-gnome on a table.<sup>13</sup> Moreover, as S\* is an internal physical duplicate of S and also has identical proximal stimulations on his/her visual receptors to S, it again seems intuitively plausible to me that the visual experiences of S and S\* have identical qualitative content.

I suppose it could be claimed here that with the above all I am doing is assuming the truth of narrow internalism with regards to the phenomenal character of visual experiences, rather than arguing for it. Indeed, it is difficult to think of an argument for supporting situation (B) which doesn't just assume some form of internalism, e.g.,

brain-in-the-vat arguments will probably only have intuitive appeal to those of a narrow internalist persuasion. The most accurate description of what S\* is going through in situation (B) is that he/she is having a type of visual hallucination, as the apparent intentional content and phenomenal character of the visual experience does not match anything in the immediate environment of S\*. However, this is no feverish or nightmarish hallucination, this is a “clear and distinct” hallucination which has the appearance of representing objects and their properties which are extended and located in space, i.e., a virtual reality.<sup>14</sup> From the subjective first-person perspective of S\*, what is “downstream” and external to the proximal stimulations on the visual receptors, cannot override what electrical pulses they send to his/her brain which then constructs the visual experience.

Briefly, there is another objection that could be made by wide externalist representationalists that S\*'s visual experience in (B) did not so much have inverted qualia, but would have different qualia, due to the different state of affairs in front of S\*, namely that there is a blue vase on a table. The upshot of this would be that the phenomenal content of S\*'s visual experience should be representing a situation where there is a blue vase on a table, rather than a garden-gnome on a table.

However, I again find this claim as implausible as the one for absent qualia, due to the fact that, by hypothesis, the proximal stimulations on the visual receptors of S\* are fixed as identical to the proximal stimulations on the visual receptors of S in situation (A). Even though the distal environment of (B) is different to that of (A), it is from the proximal stimulations in that I believe the phenomenal content of visual experiences is constructed. Therefore, in situation (B), even if light-waves may be bouncing off the distal blue vase on a table and hitting the outside surface of S\*'s

visual receptors, because the proximal stimulations on the receptors are identical to those of S in situation (A), S\*'s visual experience will have a phenomenal character that represents there being a garden-gnome on a table. Because of this, I believe that it is feasible to again consider the visual experiences of S and S\* as having identical qualitative content.

In conclusion, I have been adopting a narrow internalist representationalist stance towards the phenomenal character of visual experiences, where the latter is identified with the internal representations of the visual experience, and which are in turn locally supervenient on the internal physical properties of the subject. This seems to me the plausible position to take if one is talking about how sensory experiences seem to the subject having them. Externalism in relation to thought content has some plausibility, mainly due to compelling thought experiments by Putnam and Burge. However, for externalism to attempt to be true for how sensory experiences seem to the subject having them is, it seems to me, taking externalism too far. The proximal properties of the subject must always take precedence over the distal properties of the environment in any disputes over the qualia of sensory experiences. To put this another way, with regards to the "seemings" of sensory experience, I believe in the first-person authority of the subject. Therefore, I do not believe that the objections from wide externalist representationalism are a valid threat to the claims I have made in situation (i) from the previous chapter or to the claims for necessary/sufficient narrow conditions for belief content in this chapter.

#### 9.4 Narrow Representational Content and Phenomenal Intentionality

In this section a brief comparison of my postulated narrow representational content and the thesis of phenomenal intentionality (PI) will be made. It seems appropriate to make the comparison here as we have just been examining the notion of the phenomenal character of sensory experiences, and in particular, of visual experiences. The reason for this is that the central idea behind PI is that it is "...the intentionality a mental state exhibits purely in virtue of its phenomenal character" (Kriegel/Horgan, 2008, p.2). As we shall see, the properties that PI has are remarkably similar to the properties that my narrow representational content is supposed to have, something which I did not realise until after I had staggered and groped towards my final conception.

To illustrate the notion of PI, imagine a situation where a subject is having a visual experience of a blue vase on a table. The phenomenal character or what-it-is-like factor of this experience would be that it seems to the subject that there is a blue vase on a table in front of him/her. So, according to PI, it would be the phenomenal character of the visual experience which would fully constitute the intentionality of that visual experience (Kriegel/Horgan, 2008, p.6). Moreover, it would further be claimed that the result of this would be that there is a certain content that this visual experience contains, and that this content would be fully constituted by the phenomenology of the visual experience. We can call this content "phenomenal intentional content" (PIC) (following the terminology of Horgan/Tienson, 2002, p.524).

It is this PIC which seems to be equivalent to the narrow representational content which I have postulated. In what follows I will give several instances of the similarities between the two contents.

(i) My narrow representational content has been defined as a content that is shared between internal physical duplicates in Twin Earth-type thought experiments. The upshot of this is that the narrow content is fully constituted by the internal physical properties of the duplicates, i.e., the representational properties of the narrow content supervene locally on the subject's neurophysiological properties. Therefore, internal physical duplicates will necessarily have identical narrow representational content. Putting this another way, one could say that the narrow content shared by the duplicates is not constitutively dependent on objects and properties in the external world.

It is also the case that PIC is considered to be a content that is shared by "intrinsic duplicates" (Kriegel/Horgan, 2008, p.19). As Horgan/Tienson declare, "...phenomenology is *narrow*, in the sense that it does not depend constitutively on what's outside the skin, or indeed on what's outside the brain" (Horgan/Tienson, 2002, p.527). It is because of this that one can plausibly claim that PI and PIC are also locally supervenient on the subject's neurophysiological properties, with the result that internal physical duplicates will necessarily have, say, visual experiences with identical PIC. Therefore, it seems plausible that PIC is also a type of narrow content.



(ii) With regards to the narrow representational content of a subject's visual experiences, I characterised this as being a subjective virtual reality. It would present to the subject a virtual reality of a distal state of affairs, such as there being a blue vase on a table, which would have the appearance of being located and extended in space. The upshot of this is that the narrow content contained in the virtual reality would be considered intentional, as it was about things in the external world of the subject. It was also argued that an internal physical duplicate of the subject would experience the same subjective virtual reality, even though the immediate environment of the duplicate may be completely different, e.g., there was no blue vase on a table in front of the duplicate.

With the PI that a subject experiences, it also seems like this is a type of virtual reality of distal states of affairs. As Horgan/Tienson declare:

...sensory-phenomenal states and processes have intentional content that is inseparable from their phenomenal character. These states present an apparent world full of apparent objects that apparently instantiate a wide range of properties and relations, and they present oneself as an apparently embodied agent within that apparent world (Horgan/Tienson, 2002, p.524).

Thus it seems as if the PIC of these sensory-phenomenal states also presents the subject with a virtual reality of apparent objects and properties, as does my postulated narrow representational content. Moreover, because the sensory experiences of internal physical duplicates would be expected to share the same PIC, this would result in the duplicates also sharing the same apparent phenomenal world with its apparent objects and properties, even if those objects and properties were not in the distal environment of the duplicates. Indeed, Horgan/Tienson (2002) and

Horgan/Tienson/Graham (2004) imagine a situation where an embodied brain and a physically identical brain-in-a-vat (BIV) have the same sensory inputs throughout their respective existences. They argue that both the embodied brain and BIV would be phenomenal duplicates and so share the same PIC, despite the fact that the BIV's sensory experiences are completely illusory. This sounds similar to my strategy of attaching the duplicates to the alien supercomputer and arguing that their artificially reactivated visual experiences would result in identical subjective virtual realities of a distal state of affairs.

(iii) My postulated narrow representational content was importantly dependent on adopting the subjective first-person perspective of the person having the relevant sensory experiences and consequent beliefs. It was how the person represented things to be, or of how things seemed to the person, that was vital in formulating the narrow content.

Subjectivity is also of central importance to PI and PIC, with Kriegel/Horgan declaring that “Phenomenal intentionality is inherently subjective: when a mental state represents something phenomenally, it represents it to *someone*”

(Kriegel/Horgan, 2008, p.7). They go on to say that:

...conscious content is Janus-faced, in that in addition to its outward-looking face of presenting some object or feature or state of affairs in the world, it also possesses an inward-looking face involving an elusive *presence to the subject*...while unconscious intentional states instantiate in and of themselves only the two-place relation *x represents y*, conscious ones instantiate the three-place relation *x represents y to z* (Kriegel/Horgan, 2008, p.17).

The three-place relation is a very effective way of describing the importance of subjectivity to PI and also to my postulated narrow content. For, of course, if the narrow content of an experience or belief depends on how things seem, then there always has to be a conscious subject that is having the seeming to. One of the points I made against externalism with regards to mental content was that it generally downplayed the importance of how things seem to the person in determining the content that their particular experience or belief had. For externalism, it seemed more important what object or kind was actually in the person's environment, even if the person concerned had no knowledge of this, in determining the content of the person's experience/belief. I called this tendency adopting the third-person objective viewpoint when ascribing mental content. Hopefully, I argued persuasively that it was far more plausible to sometimes adopt the first-person subjective viewpoint of the person having the experience/belief when ascribing a certain content to it.

(iv) With the postulated narrow content, I claimed that it possessed significant truth-conditional and referential properties. The narrow truth conditions of a belief would only be determined from the experiential and conceptual properties involved with the belief. For example, the narrow truth conditions of the belief that water is wet are that it would be true if and only if the transparent, colourless, odourless, liquid that comes out of taps, falls from the sky as rain, etc., is wet. What I called the wide truth conditions of the belief that water is wet would involve the microstructure of the water, so that on Earth the truth conditions would be that H<sub>2</sub>O is wet and on Twin Earth the truth conditions would be that XYZ is wet. Concerning the referential properties of the narrow content, I thought that this could plausibly take an intensional form, with the intension of , say, a term, determining what distal object or state of

affairs would be validly contained in the extension of that term. The validity would be determined by the intension by again using something like the experiential and conceptual properties involved with the term, where the distal object or state of affairs that was correctly referred to by the term would have to possess certain properties. For example, with the term “water”, this would narrowly refer only to the transparent, colourless, odourless, liquid that comes out of taps, falls from the sky as rain etc., in the surrounding environment.

The notions of PI and PIC advocated by Horgan/Tienson (2002) and Horgan/Tienson/Graham (2004) also involve narrow truth conditions.

Horgan/Tienson put it like this:

Suppose Alfred and Bertrand are looking at two different barns, and each of them says, “That’s an old barn.” Do their statements have the same truth conditions? Yes and no. In one way, they have different truth conditions. Alfred’s statement is made true or false by the age of the barn that he is looking at, while Bertrand’s statement is made true or false by the age of the distinct barn that *he* is looking at...we will call such truth conditions, which depend on the actual entities referred to in a statement or thought, *wide truth conditions*. But in another way, Alfred’s and Bertrand’s statements have the same truth conditions. In each case the truth condition is that there must be an actual barn that he is looking at...and that barn must be old. Such truth conditions are *narrow truth conditions*. They are determined skin-in, so to speak, and are completely determined by phenomenology (Horgan/Tienson, 2002, p.528).

Therefore, it can be seen that the role of phenomenology in determining the narrow truth conditions of the thoughts is equivalent to the role that I gave to experiential and conceptual factors in determining the narrow truth conditions of the water-beliefs. Moreover, like the experiential/conceptual factors, the phenomenology is considered to be determined only by properties internal to the subject. As with the wide truth conditions for the water-beliefs I mentioned previously, the wide truth conditions for

Alfred's and Bertrand's thoughts depend on a specific object (or kind) being present in their respective environments, i.e., the specific barns they are looking at.

It also seems that the PIC of Horgan/Tienson (2002) and Horgan/Tienson/Graham (2004) also contains something similar to an intension which gives the content its referential properties. Horgan/Tienson declare that the PIC of a thought can provide what they call "grounding presuppositions", which will effectively provide the properties that an object or individual in the world would have to possess in order for the thought to correctly refer to that object or individual (Horgan/Tienson, 2002, p.528). The situation now gets slightly complicated, but if I understand things correctly, the grounding presuppositions themselves are completely determined by the phenomenology of the thought, and so are considered to be narrow (Horgan/Tienson, 2002, p.529; Kriegel/Horgan, 2008, p.34). It is the narrow and phenomenologically determined grounding presuppositions which originally determine which object or individual is referred to by the thought in question. At this stage, the grounding presuppositions of the thought are "*intentionally directed toward*", or "*purporting to refer to*" an object or individual in the external world (Horgan/Tienson, 2002, p.529).

So far, the grounding presuppositions of a thought seem analogous to the intension of a thought that is provided by the postulated narrow representational content of that thought. Both entities narrowly determine what the thought would correctly refer to in the external world. However, there now enters what seems to be a disanalogy, as Horgan/Tienson claim that if there is an actual, existing, "satisfier" of the thought's grounding presuppositions in the world, the thought then *refers* to that satisfier, but that this will result in the thought having wide content and wide truth conditions

(Horgan/Tienson, 2002, pp.528-529). But this does not seem correct for the way that I understand the intension, provided by the narrow representational content of a thought, to operate.

To illustrate this, it will perhaps be useful to consider an example.<sup>15</sup> Imagine that I have the thought that water is wet on Earth and my doppelganger has the thought that water is wet on Twin Earth, with the usual differences between the waters. According to Horgan/Tienson, the PIC of me and my doppelganger's water-thoughts will be identical, which will result in identical grounding presuppositions for the reference of those thoughts. As these grounding presuppositions are fully determined by the thoughts' phenomenology, they are considered to be narrow. I will assume that the grounding presuppositions for the respective waters are something like - the transparent, colourless, odourless, liquid, which comes from taps, falls from the sky as rain, etc. In this sense, the grounding presuppositions, although they are phenomenologically determined and narrow, are intentional or world-directed. However, once there are actual, existing, satisfiers for the grounding presuppositions of the water-thoughts, the latter then correctly refer to those satisfiers, with the result that the water-thoughts now have wide content, i.e., the wide content of my water-thought on Earth will be H<sub>2</sub>O and the wide content of my doppelganger's water-thought on Twin Earth will be XYZ.

This then is how I understand Horgan/Tienson's reasoning on the grounding presuppositions involved with the PIC of a thought. However, this is not how the intension produced from the narrow representational content of a thought would behave. The intensions for me and my twin's water-thoughts would be determined by

the experiential and conceptual factors involved with the thought, and would be considered type-identical, as we are internal physical duplicates. The result of this would be that the intension for our respective “waters” would be something like the colourless, odourless, liquid which comes from taps, falls from the sky as rain, etc. So far, me and my twin’s narrow intensions would be the same as me and my twin’s grounding presuppositions.

However, when the intension for my water-thought determines the reference for it, and it denotes the “watery-stuff” in my environment, it is not denoting the wide content of that water, i.e., H<sub>2</sub>O, it is just denoting the watery-stuff, and the same, *mutatis mutandis*, for my twin on Twin Earth. The upshot of this is that the intensions of me and my twin’s water-thoughts not only have narrow type-identical intentionality, but they also have narrow type-identical reference. Of course, in terms of the wide content of the water-thoughts, the intensions have type-distinct reference, but there is no wide content involved with the intensions, they have been determined only from the narrow representational content of the subject’s experience.

Therefore, the shared narrow intension of me and my twin’s water-thoughts would pick out water on the Earth and Twin water on Twin Earth, and the water-thoughts would be referring, but only referring to the narrowly construed, distal, watery-stuff in our respective environments. By this, I mean that even though the watery-stuff is in the external world of the subject, when a subject’s water-thought narrowly refers to it, the content does not become wide, but stays as narrow content. This then seems to be the main difference between the intensions that are formed from narrow representational content and the grounding presuppositions that are formed from PIC.

However, it should be said that it only seems to be Horgan/Tienson's (2002) and Horgan/Tienson/Graham's (2004) version of PI that involves wide content with its mental reference. Other versions of PI seem to argue for positions closer to my own regarding their referential properties. For instance, Kriegel (2007, 2008) tries arguing that the PI of a thought about, say, Bigfoot, is the non-relational property that constitutively determines that the thought is representing Bigfoot (Kriegel, 2007, p.317). Moreover, Kriegel also portrays having a thought about Bigfoot in an adverbial manner, so that it comes out as "you are thinking Bigfoot-wise" (Kriegel, 2007, p.314).<sup>16</sup> The upshot of this is that when one is having a water-thought one would be non-relationally representing water-wise, where the PI of the thought would provide accuracy conditions which would have to be satisfied in order for the thought to validly refer to something in the external world (Kriegel, 2008, p.90). Also, Farkas (2008) tries arguing that the PI of the thoughts of twins could result in the content of those thoughts staying the same, even though the reference of the thoughts are different (Farkas, 2008, p.289). This result comes about because Farkas claims that "...the difference in reference is due to factors external to this content" (Farkas, 2008, p.289).<sup>17</sup>

Therefore, this has been a brief comparison between my postulated narrow representational content and the notion of PI. It has been seen that there are several similarities between the two notions, which perhaps, is not that surprising, as the determination of the mental content concerned with both notions does rely heavily on how things seem to the subject as he/she is having the relevant experience or thought.



## 9.5 Necessary/Sufficient Narrow Content and Explanations of Intentional Behaviour

It might be thought that the above conclusions supporting the claim that narrow representational content provides necessary and sufficient conditions for certain belief-contents could also accommodate the arguments of Noonan (1986, 1991, 1993) against the claim of Evans (1982) that singular object-dependent thoughts (or Russellian thoughts) are required for a satisfactory psychological explanation of a subject's intentional actions.<sup>18</sup> The definition of a singular object-dependent thought is here considered to be one which contains a demonstrative or proper name, and concerns a particular object that is external to the subject, an object whose existence is essential to the possibility of having the thought in question, i.e., if there is no object, then it is not possible to have a thought about it. Noonan's argument against Evans' claim involves supporting what he calls Thesis R:

Whenever an action is directed towards a concrete, contingently existing object, other than its agent, in the sense that it is intentional under a description in which there occurs a singular term denoting that object, then an adequate internalist [object-independent] psychological explanation of it is available under a (possibly distinct) description in which occurs a term denoting that object (Noonan, 1991, p.2).

The argument can be briefly illustrated as follows. Imagine that S is in situation (A), having the veridical experience of perceiving the garden-gnome on a table, i.e., it is a veridical situation because S's internal representations of the situation actually have their external world referents. With this in mind, imagine that S points at the garden-gnome and exclaims "That garden-gnome is from Argos!" Now imagine that S is attached to the alien supercomputer and is having an organismically formulated

virtual experience of situation (A), where from S's point of view everything appears exactly the same as in the veridical situation. Basically, what is happening here is that when attached to the alien supercomputer, S is only experiencing his narrow representational content of the situation of the garden-gnome being on the table, there are no external world referents to the internal representations. Nevertheless, in this nonveridical situation, S again points at the [garden-gnome] and exclaims "That garden-gnome is from Argos!"

In a nutshell, what Noonan claims from this is that an adequate psychological explanation of S's actions in the nonveridical situation can be contained in a subset X of S's psychological states in the veridical situation. However, if it is then assumed that the contents of S's psychological states (contained in subset X) are the same in the veridical and nonveridical situations, then subset X could also give an adequate (non-relational) psychological explanation of S's actions in the veridical situation, the only difference being a relational one, i.e., in the veridical situation the garden-gnome actually exists. The upshot of this is that object-dependent thoughts about the garden-gnome seem to be psychologically redundant in explaining S's actions in the veridical and nonveridical situations (Noonan, 1991, pp.2-5).

However, the interesting part of all this is that, if I understand Noonan correctly in his 1993 article, he would not accept the above illustration of his argument using the alien-simulated virtual experience of S. I come to this conclusion because Noonan is of the opinion that his argument for Thesis R would not work with an example that involved the psychological states of a brain-in-a-vat (BIV) and the psychological

states of an embodied subject (Noonan, 1993, p.295). That is, Noonan believes that

we cannot:

...think of envatted HN [Harold Noonan] as *intentionally performing actions which are identical with embodied HN's actions (non-relationally described), though having available no content or component of content unavailable to embodied HN* (Noonan, 1993, p.295).

He gives three reasons for holding the above conclusion, none of which strike me as

particularly compelling. Firstly, Noonan claims that:

...it is not obvious that we can regard envatted HN as possessing any psychological states at all (so that his difference from embodied HN might merely come down to the fact that the latter possess, whilst he lacks, a mental life) (Noonan, 1993, p.295).

Now, why would we doubt that a BIV had psychological states, or indeed that the poor thing even had a mental life? Rather, it seems intuitive to say that having a mental life is about the only thing that a BIV has going for it. Surely the identical brain states of envatted HN and embodied HN result in some sort of narrow mental life that is shared between them, at least if we consider that our (narrow) psychological states have some sort of supervenience relationship to our brain states.

Secondly, Noonan states that:

...even if we ought to regard envatted HN as possessing psychological states, it is not evident that we can regard him as intentionally performing any actions which are identical with embodied HN's actions (non-relationally described)...it is not obvious how we can ascribe non-observational knowledge of any such [intentional] actions to a brain in a vat. For we do not, as embodied beings, have non-observational knowledge of the activities of our own brains (non-relationally described) (Noonan, 1993, p.295).

What Noonan is alluding to here is that in the example that he uses for supporting Thesis R, there are two embodied HNs, who both make the same bodily movements (of kicking a cat), the only difference being that one HN actually kicks a cat whilst the other is only hallucinating a cat (Noonan, 1991, p.3). So deluded HN has non-observational knowledge of “lashing out with his foot”, which a BIV would not have. However, it seems to me that Noonan is looking at this situation from the wrong perspective, i.e., the objective third-person perspective. It may be the subjective first-person perspective of the BIV that we need to take into consideration here, when making judgements about intentionally performed actions by the BIV (indeed, in the original example, Noonan declares “Imagine now a second situation in which, *from my point of view*, everything is the same, but in which, in fact, I am hallucinating a cat” (Noonan, 1991, p.3) [my italics]. For instance, imagine that the BIV is connected up in a similar manner to the way that S is connected up to the alien supercomputer, with a screen that shows a simulation of the BIV’s subjective perceptions and experiences from its first-person perspective. On the screen we would see a cat and the BIV’s virtual leg kicking out at it, we could also imagine that the BIV would subjectively feel and see his leg kicking out at the cat. In such a situation as this, it seems more plausible to consider the BIV as intentionally performing actions which are identical to those of the embodied HN, at least when viewed from their respective subjective first-person perspectives.

Thirdly, Noonan declares that:

...it is not obvious why, if we can ascribe contentful psychological states to envatted HN, we cannot regard him as possessing contentful psychological states with contents not available to embodied HN (as Putnam (1981) puts it, not contents concerning brains and vats, but rather concerning brains and vats

‘in-the-image’). For, of course, the external relations of his brain-states are quite different from those of embodied HN (Noonan, 1993, p.295).

However, this seems to me a rather strange objection for Noonan to make, as he himself admits, this line of reasoning could also have been taken against his argument for Thesis R, i.e., in the case of kicking the cat, it could be claimed that the psychological states of deluded HN had a different content to the psychological states of non-deluded HN, due to the fact that the former’s brain-states had different external relations to the brain-states of the latter. Non-deluded HN’s brain-states were externally related to a cat, whilst deluded HN’s brain-states were not. However, in response to this, all Noonan says is that “...in this [cat-kicking] case the claim would be completely implausible” (Noonan, 1993, p.296).

But why should this objection be implausible in the cat-kicking example and plausible in the case of envatted HN? There is a difference in the external relations between the brain-states of non-deluded HN and deluded HN, but according to Noonan the contents of their respective psychological states contained in subset X will be the same (Noonan, 1993, p.286). It seems intuitively plausible to me that it would also be the same story with embodied HN and envatted HN. That is, the contents of envatted HN’s psychological states would be a subset X of the contents of embodied HN’s psychological states. In support of this claim I can also point to Putnam (1981), where he accepts that the notional world of a BIV would be identical to the notional world of its embodied twin-brain (Putnam, 1981, pp.28-29); or put another way, Putnam claims that the bracketed or pure thoughts of a BIV would be qualitatively identical to the bracketed or pure thoughts of its embodied twin-brain (Putnam, 1981, pp.42-43). For Putnam, bracketed or pure thoughts are thoughts which do not have

external world reference, and so they seem to be equivalent to the thoughts contained in subset X, as these have non-relational specifications.

Therefore, if we now follow Noonan's argument through as before, the subset X of contentful psychological states of the BIV could give an adequate psychological explanation of embodied HN's actions (non-relationally described). If this is accepted, then the conclusion is that if embodied HN does have any object-dependent thoughts (which of course, will not be contained in subset X), these are psychologically redundant in giving an adequate explanation of embodied HN's actions.

The upshot of all this is that, contra Noonan, his argument against the claim that object-dependent thoughts are required in order to give adequate psychological explanations of a subject's intentional actions, can also be used in the situation of a BIV and embodied subject, and so, ipso facto, the argument could also be used with the veridical situation of S pointing at the garden-gnome and the alien-simulated nonveridical situation of S's virtual arm pointing at the virtual garden-gnome.

Moreover, if it is accepted that Noonan's argument for the object-independence of singular thoughts in psychological explanations can be applied to the situation of a BIV and embodied subject, then once again, contra Noonan, his argument could be utilised as an argument against externalism in general (Noonan, 1993, p.294).

The above then is my attempt at arguing that the postulated narrow representational content provides necessary and sufficient conditions for certain beliefs to have the contents that they do. I am not sure how successful or plausible this attempt has been,

but by taking into account the most powerful objections to this narrow content and trying to show how they can be overcome, I have tried to make the claim for it as positive and powerful as I can.

## Notes to Chapter 9

<sup>1</sup> It should be pointed out here that all the belief-contents I refer to in this chapter are essentially *perceptual* belief-contents.

<sup>2</sup> I should also say here that I had the idea for a scenario like the above from McGinn, 1989, pp.58-99, where he is trying to ascertain whether perceptual content has strong or weak externalist properties. In trying to answer this question, McGinn imagines a situation where a subject has experiences as of square things, even though the distal causes of those experiences are round things.

<sup>3</sup> Following Lewis (1980), perhaps we could say here that in situation (B), S\* is not perceiving or seeing the scene before him. Lewis defines seeing thus:

...if the scene before the eyes causes matching visual experience as part of a suitable pattern of counterfactual dependence, then the subject sees; if the scene before the eyes causes matching visual experience without a suitable pattern of counterfactual dependence, then the subject does not see (1980, p.87).

What I think Lewis means by this is that a subject sees if in the actual situation, the scene before him causes a matching visual experience, but also in a counterfactual situation where there is a different scene facing the subject, this different scene causes a different visual experience which matches the counterfactual scene. However, does this mean that with situations (A) and (B) above, S would not be seeing in (A), due to the fact that S\*'s counterfactual visual experience does not change to match the different scene? That is, for S to be correctly considered to be seeing the garden-gnome on the table in (A), S\*'s visual experience should have changed to that of a blue vase on the table due to the fact that the counterfactual scene facing him is of a blue vase on the table?

I am not sure about this. It seems to me that in (A), it would be intuitive to say that S is seeing the garden-gnome on the table, as his visual experience is (albeit contingently) caused by, and matches, the scene facing him, i.e., a garden-gnome on a table. Then in counterfactual situation (B), S\* would be judged not to be seeing the scene before him, as his visual experience has not changed to match the different scene before him, i.e., a blue vase on a table. Moreover, even if in (B) the scene facing S\* had been identical to that in (A), that is, a garden-gnome on a table, we should still not say that S\* is seeing a garden-gnome on the table, as S\*'s visual experience of a garden-gnome on the table had not been caused by the counterfactual garden-gnome on a table, his visual experience of a garden-gnome on the table had been frozen from the actual situation. What we would have in a situation like this would be what Lewis calls a *veridical hallucination*, where the visual experience matches the scene before the eyes, but is not classed as seeing because the visual experience is not causally dependent on that scene. As situation (B) is set up, perhaps this could be called a nonveridical hallucination, as the visual experience of S\* is not causally dependent on the scene before his eyes, but that visual experience does not match the scene anyway.

<sup>4</sup> What I am presupposing here is that there is a necessary connection between certain proximal stimulations on a subject's visual system and the resultant neural representation that is constructed; whilst I previously claimed that there was not a necessary connection between certain distal objects causing certain proximal stimulations on a subject's visual system. It is possible that the former presupposition could be questioned, although I think it would be difficult to construct a counterexample as the subjects concerned are intrinsic physical/functional duplicates.

<sup>5</sup> Of course, a subject's narrow belief-content which has been produced from his internal representation of a certain state of affairs may well turn out to be false, if the state of affairs being internally represented is not actually present in his immediate external environment. However, I don't think this would affect the above intuition that his narrow belief has the particular content that it has due to his internal representations.

<sup>6</sup> It should be noted that McGinn does not actually use a counterfactual scenario in his Percy thought experiment, but follows Putnam by using the Earth and Twin Earth as different planets in the same actual universe. I don't think much turns on this point, and I am basically following Davies' description of McGinn's thought experiment, which does use a counterfactual scenario (1991b, pp.35-36).

<sup>7</sup> Using Davies' terminology, this conclusion is a "conservative individualist" stance towards McGinn's thought experiment, as I am claiming that the counterfactual perceptual content of internal state S1 is the same as the actual perceptual content of internal state S1. This is in contrast to a "revisionary individualist" stance, which would dispute the claim that the actual perceptual content of an internal state is what it is, typically claiming that it should be a more inclusive and disjunctive perceptual content. It is then claimed that the counterfactual perceptual content of that internal state has the same inclusive and disjunctive content as the actual internal state. See Davies, 1991b, pp.28-31 for a more thorough explanation of these different individualist stances.

<sup>8</sup> McGinn's weak externalism seems similar to Recanati's (1994) notion of a "relative" narrow belief content, which is considered to be independent of the subject's actual environment, but is still dependent on the subject's normal environment (1994, p.221). That is, counterfactual-Percy could still have a perceptual content as of a square thing, even though there are no distal square things in his immediate environment when he is having the perception. However, his perceptual content would still be externalist in nature, as in his normal environment, his having a perceptual content as of a square thing would depend on there being a distal square thing present. Recanati's aim is to construct a narrow content which can be part of a two-component theory of content which is compatible with radical externalism.

<sup>9</sup> This would also be my response to Edwards (1994), who criticises McGinn's "Percy" thought experiment as part of a defence of Burge's (1986, 1988a) cracks/shadows argument for the externalism of perceptual content. Using the "Percy" thought experiment as set up in this section, Edwards questions the plausibility of ascribing a perceptual content as of a square thing to counterfactual-Percy, when there are no distal square objects in the counterfactual scenario (and so counterfactual-Percy would also not have the concept "square" (1994, pp.81-84). However, if we imagine the Percys sharing a narrow subjective virtual reality which contains an instance of a square thing, i.e., it seems to counterfactual-Percy that he is having a visual experience of a square thing, then the possibility of ascribing to him a perceptual content as of a square thing becomes more plausible. Moreover, he would have the concept "square" because the Percys are internal physical/functional duplicates and actual-Percy possessed the concept "square". Therefore, contra Edwards, despite the fact that there are no instances of distal square objects in the counterfactual scenario, it still seems plausible to ascribe the perceptual content as of a square thing to counterfactual-Percy; which also makes it more plausible that he is continually misrepresenting distal round things as square things.

<sup>10</sup> Support for wide externalist representationalism can be found in Dretske (1995), Lycan (1996), and Tye (1995, 1998, 2000).

<sup>11</sup> Gray (2003) argues against an absent qualia claim contained in Tye (1995, 2000). The absent qualia claim comes from Tye's so-called PANIC theory for deciding which perceptual states are correctly called conscious ones and from his causal co-variation theory of intentional content (2003, p.246). The strange upshot from all this is that Tye argues that a simple creature which has perceptual states related to a single content can be said to be having conscious experiences, whereas, a simple creature which has perceptual states related to a disjunction of contents cannot be said to be having conscious experiences. Gray argues persuasively against this by providing a counterexample involving a creature which has perceptual states which are related to a disjunction of heat contents, i.e., kinetic energy and radiant energy. Despite the perceptual states having a disjunctive content, it is intuitively plausible that the perceptual states of the creature will still be conscious ones (2003, pp.248-250).

<sup>12</sup> Indeed, the connection between the phenomenal character and intentional content of conscious experiences may be very close. Both Siewert (1998) and Horgan/Tienson (2002) have argued that it is very difficult to conceive of a situation where two conscious mental states have identical phenomenal character but have a significant difference between their intentional contents.

<sup>13</sup> To put this another way, it seems plausible that things can look such-and-such to S\*, even though there are no objective physical surfaces being represented with the visual experience. This way of putting things is taken from Alston (2005), where he is talking about the basic visual experiences of frogs and insects (2005, pp.285-286). The target of Alston's article is also externalist representationalism, although his preferred theory of perception is a strong version of direct realism called the "Theory of Appearing" (2005, pp.256-257).



<sup>14</sup> Has there ever been such a powerful and convincing hallucination in real life as the one postulated in situation (B)? To be honest, probably not. Indeed, Putnam (1999) claims that “It may well be that in actual life, as opposed to science fiction...a dream or hallucination that is “qualitatively identical” (or indistinguishable from) a waking/normal experience simply never occurs” (1999, p.153). As Putnam says, perhaps it’s better not to rely on intuitions that come from science fiction scenarios, especially ones that posit identical planets to Earth, doppelgangers, XYZ, and so on. However, Carter (2000) seems to give support to Putnam’s doubts about the qualitative power of an hallucination where she claims that the difference between actual seeing and hallucinating is that in the former case “...more sensory neurons are activated in response to outside stimuli than in self-generated sensory experience” (2000, p.206). But she goes on to say that people with photographic or eidetic memory can “...create visualizations that are just as intense as those brought about by the original stimuli” (2000, p.206). Therefore, this seems to indicate that it is sometimes possible for visual hallucinations to be qualitatively identical to actual seeing, particularly if the amount of self-generated neurons firing in the subject’s visual system is the same as the externally-stimulated amount of neurons firing during actual seeing. Carter gives a number of examples of visual hallucinations, which appear to be clear and distinct to the people having them, e.g., visual hallucinations of large buildings, familiar people, crowds of people, even a herd of cows (2000, pp.203-204). All of this seems to indicate that it is perhaps not so implausible for S\* in situation (B) to have a visual experience of a garden-gnome on a table, particularly if the proximal stimulations on his/her visual receptors result in an identical amount of neurons firing in the visual cortex as would a veridical perception of a garden-gnome on a table.

<sup>15</sup> Horgan/Tienson use an example about an Earth-thought referring to Bill Clinton and a Twin Earth-thought referring to the Twin Earth Bill Clinton, but they do claim that reference to natural kinds would behave similarly (2002, p.529), so I will stick with water and Twin-water for my example.

<sup>16</sup> Kriegel claims that the outcome of these moves is that one can think about Bigfoot without needing a constitutive representation relation between the thought and what the thought is about, i.e., Bigfoot (2007, p.315). For Kriegel, this solves the problem of intentional inexistence, i.e., being able to have thoughts about non-existents.

<sup>17</sup> With regards to thoughts and their contents, Farkas goes on to claim that “...it seems equally clear to me that in certain third-person mental state attributions, there is a way of talking about the objects of someone’s thoughts which does not shed light on those thoughts’ nature” (2008, p.290); and that “...when we describe what someone thought or said we make use of information which is available to us, but not necessarily available to the subject. This is a perfectly legitimate procedure, but may involve elements which are not relevant to the nature of the subjects’ thoughts *qua* thoughts” (2008, p.290). All I can say to this is that I agree wholeheartedly!

<sup>18</sup> It is worth noting here that Segal (1989a) uses an argument similar to Noonan’s to argue against McDowell’s (1984, 1986) claims for object-dependent thoughts, or, to be more precise, object-dependent Fregean senses.

# MAIN CONCLUSION

In this final part of the thesis I will try to state what conclusions can be drawn about the notion of narrow content in the philosophy of mind. This will be done by looking back over the four accounts of narrow content that have been examined, particularly the last account which I proposed. It will then be decided whether an adequate and plausible narrow content has been produced, or if, to mis-paraphrase Wittgenstein, I have just been trying to help an elderly fly back into its former abode, i.e., the bottle.

The definition of narrow content proposed here is that it is a type of mental state that is shared between doppelgangers or internal physical duplicates in Twin Earth-type thought experiments. To gauge whether a narrow content construal is adequate, I introduced three conditions of adequacy which the construal would have to satisfy in order to count as adequate. These conditions are: (1) The narrow content construal must make it plausible that there is such a thing as an internal, “in the head”, type of mental content, that is shared by duplicates in Twin Earth-type thought experiments. (2) The narrow content construal must be such that a plausible specification of that content can be given. (3) The narrow content construal must provide an adequate psychological generalisation that explains the behaviour of both duplicates.

Additionally, the construal must be flexible enough to make more fine-grained distinctions in mental content, e.g., when dealing with Frege cases.

Firstly, Loar's notion of psychological content was taken as the exemplar of a conceptual role construal of narrow content. The psychological content of a belief consisted in the way that the belief conceptually interacted with the subject's other beliefs, sensory inputs and behavioural outputs. In relation to Burge's "arthritis" thought experiment, Loar argued persuasively that the psychological content of actual-Bert's and counterfactual-Bert's arthritis beliefs were the same, even though it might have been correct to ascribe distinct de dicto arthritis beliefs to them.

However, the problem with a conceptual role construal of narrow content is that giving a determinate specification of that narrow content is difficult, due to the fact that the narrow belief content is determined solely by its conceptual relations with other beliefs, etc. It was because of this that I felt Loar's psychological content did not satisfy the adequacy conditions (2) and (3) and so was not an adequate narrow content.

Nevertheless, during this consideration of Loar's psychological content, an idea was floated whereby objective and subjective perspectives could be adopted when ascribing beliefs. The upshot of this was that when beliefs were ascribed from the subjective perspective, they could be given a subjective de dicto specification, which took account of how the subject was conceiving things, which might sometimes be overlooked when ascribing beliefs from the objective perspective. There seemed the possibility that the narrow content two beliefs shared could be specified using subjective de dicto belief ascriptions.

Next, Fodor's mapping theory was examined, which was taken as an exemplar of an indexical construal of narrow content. The mapping theory took as its inspiration Kaplan's notion of the character of an indexical expression, which Kaplan suggested acted as a function from contexts of utterance to the content that the indexical expression would have. However, Fodor's notion of character, which he identified with narrow content, would be involved with mental states, and was defined as being a function from contexts of thought to wide contents. The narrow content of the thoughts would map those thoughts onto the same wide contents in the same context. For example, the narrow content that is shared between me and my duplicate's "water" beliefs would act as a function which mapped my "water" onto XYZ in my duplicate's Twin Earth context, and would map my duplicate's "water" onto H<sub>2</sub>O in my Earth context. Therefore, when me and my duplicate are in the same environmental context, our "water" beliefs have the same wide content.

The problem for Fodor's mapping theory was to specify what the narrow content of the beliefs was, which explained how it behaved like a function from contexts to wide contents. It seemed natural to construe the narrow content or character of the beliefs as an intension, which would determine an extension, due to the imposition of extension conditions on those beliefs. For example, for some external world object to be in the extension of a "water" belief, it would have to satisfy certain conditions, which could involve conceptual and observable properties, like being a transparent, odourless, colourless liquid, which fills the oceans, falls from the sky as rain, etc. The idea of an intension for a belief seemed a useful way of getting out into the external world in a controllable manner.

However, the mapping theory did have a number of drawbacks. It could still be doubted whether it did provide a plausible internal belief content that both duplicates shared, as Fodor was of the opinion that the beliefs only gained semantic content when anchored in a context, i.e., when the beliefs had wide content. In relation to this point, it seemed that any psychological generalisation to explain behaviour that the theory provided, would again only involve wide content (which would also mean that it could not deal with Frege cases). It was for these reasons that I thought the mapping theory failed to satisfy adequacy conditions (1) and (3) and so was not an adequate narrow content.

The third construal of narrow content was then examined, that of narrow representational content, the exemplar of which was Dennett's conception of notional worlds. I'm not sure that Dennett himself would have agreed with his notional worlds being construed in a narrow representationalist manner, but the idea of it seemed very plausible to me. Dennett's original idea of how the notional world of a subject was to be formulated seemed to me rather implausible, as it depended on what Dennett called hetero-phenomenology. This involved an observer formulating the subject's notional world from the third-person perspective, mainly utilising the subject's behavioural and verbal dispositions. It was because of this that I felt the original notional worlds conception did not provide a plausible narrow belief content which was an internal or "in the head" content shared between duplicates. For this reason it did not satisfy adequacy condition (1) and so was not an adequate narrow content.

It seemed to me that a more intuitive strategy was for a subject's notional world to be formulated from the first-person perspective of the subject, involving how the subject

represented things to be. This in turn involved the internal representations of the subject's environment, but these representations would be determined from the proximal stimulations in, thus making them, as far as I was concerned, to be internal states of the subject. The upshot of this was that two internal physical duplicates on Earth and Twin Earth could have identical notional worlds involving a certain state of affairs, such as seeming to have a perceptual experience of a garden-gnome with water coming from his bottom. The consequent beliefs they would have could also be given a narrow specification in a subjective de dicto belief ascription, where the term "water" would have the shared meaning of being the transparent, odourless, potable, etc., liquid in the environment, i.e., the meaning would involve what Putnam called the stereotype of water. The different chemical compositions of the respective waters (H<sub>2</sub>O and XYZ) would be irrelevant to the semantics of the narrow subjective de dicto water-beliefs of the duplicates.

The point in the thesis had then been reached where I had to put forward my own preferred construal of narrow content. I decided on a construal similar to my narrow representational version of Dennett's notional worlds. Once again considering the duplicates' internal representations to be constructed from the proximal stimulations on the sensory receptors of the duplicates, it struck me that the internal representations could form a narrow virtual world, at least from the subjective point of view of the subject. For example, if one considered the visual perception of a certain state of affairs, such as there being a glass of water on a table, a virtual reconstruction of this state of affairs could occur from the proximal stimulations in. In terms of visual phenomenology, the virtual state of affairs would be identical to the veridical perception of the distal state of affairs, i.e., from the first person perspective of the

subject, it would seem that he/she was having a visual experience as of a glass of water on a table in both the veridical and nonveridical situation. The claim was made that these subjective virtual realities would be shared between the duplicates, with the result that their narrow representational content was identical.

From this, I proceeded to argue that this narrow representational content did possess considerable semantic properties involving meaning and reference. Effectively, what I claimed was that the semantic properties that were contained in a virtual reality of a certain situation were very similar to the semantic properties that were contained in a veridical perception of that state of affairs. The only difference would be that with the distal state of affairs, one would be able to reach a further semantic distance into the world and differentiate the different chemical compositions of the respective waters (i.e., H<sub>2</sub>O and XYZ). Nevertheless, the upshot of this was that the narrow representational content shared by the duplicates could once again be given a plausible specification using subjective de dicto beliefs, where the meaning and reference of “water” would be based on the visible and conceptual properties associated with water.

Finally, to make the postulated narrow representational content as positive and powerful as I could, an attempt was made to argue that this narrow content could provide necessary and sufficient conditions for having a belief with a certain content. My strategy involved a thought experiment which kept the proximal stimulations on the visual system of the duplicate the same, in the actual and counterfactual situations. The result of this, so I argued, would be that the counterfactual duplicate would have an internal representational experience identical to the one that he/she had in the

actual situation, even though the distal state of affairs in front of the duplicate in the counterfactual situation was different to that of the actual situation. It was then argued that these identical narrow experiences would lead to the formation of beliefs with a certain content.

Moreover, I believe that my proposed narrow representational content is an adequate narrow content as it can satisfy all three conditions of adequacy, something which the other construals could not do. It satisfies condition (1), positing a narrow belief content that is internal to, or “in the head” of, both duplicates in a Twin Earth-type scenario. It does this by making the narrow representational content locally supervenient on the neurophysiology of the duplicates. My narrow content also satisfies condition (2), in that it is capable of being given a plausible specification using that-clauses, albeit the latter are of the subjective *de dicto* type. Finally, my narrow content satisfies condition (3), in that it provides a plausible psychological generalisation that subsumes the behaviour of both duplicates in a Twin Earth-type scenario. Moreover, it also seems plausible that my narrow content construal could deal effectively with so-called Frege cases, as the representational character of it tries to capture how the subject is representing, or conceiving of, things in the world.

Therefore, I believe that the narrow representational content that I have postulated is an adequate and plausible narrow content. In putting forward my account of narrow content, I concentrated on Twin-Earth situations involving natural kinds and differences in the physical environment of the subject, as it is in this area that I think the externalist claims for mental content are the strongest. However, it is my opinion that the narrow representational content postulated could also deal effectively with



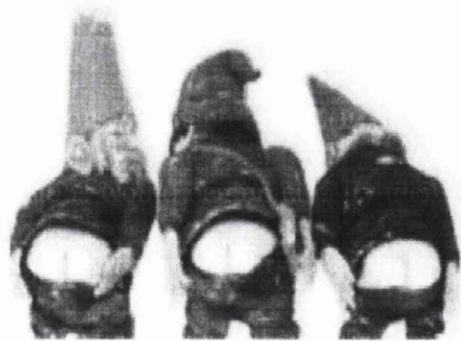
Burge-like situations involving differences in the social environment of the subject. After all, the narrow representational content is supposed to capture how the subject conceives of things from his/her own first-person perspective. It seems plausible to imagine that with Burge's "arthritis" thought experiment, this construal could also capture a narrow content associated with the term "arthritis" that is shared between duplicates from their respective subjective viewpoints.

Indeed, an important corollary of this is that if Twin Earth-type thought experiments are intended to refute internalism about mental content, i.e., by claiming that even though the internal physical properties of the duplicates are kept identical, the contents of their mental states have still changed due to differences in their environments, then the existence of a plausible mental content that is shared between the duplicates would seem to indicate that internalism has not been completely refuted.

This then brings me to the end of the thesis. Hopefully, to refer back to Wittgenstein's analogy of the fly in the bottle, I have convinced you that I haven't helped the fly back into the bottle with respect to the notion of narrow content, and that there is still some life left in the internalist position on mental content, even though it's not very popular at the moment. The key to a plausible narrow content seems to be taking into account the subjective point of view of the person having the belief, when making belief ascriptions. It is how things seem to the subject, how he/she represents things as being, that is important in giving a narrow reading of their beliefs and other attitudes. This is sometimes in contrast to belief ascription from the objective point of view, which can move away from the perspective of the subject,

and which, although it probably includes accurate objective information, can be inaccurate in capturing how the subject conceives things as being.

If I can use an analogy myself to describe the situation, perhaps having beliefs and other attitudes from the first-person perspective is like our perception and knowledge of the constellations in the night sky. From the perspective of the Earth, we can make out the shapes that the stars form, such as Ursa Major or The Great Bear. However, if we change the perspective from which we view Ursa Major, perhaps by travelling a couple of hundred light years away from the Earth, we could view the same stars that make up Ursa Major from the Earth-perspective, but from this new perspective they would have a different formation and shape. From an objective point of view, the stars involved on both occasions would be numerically identical, and I suppose this is interesting knowledge to have about the universe, however, it is only from the Earth-perspective that those stars have the shape and signification that they do.



**THE END(S)**

# BIBLIOGRAPHY

- Adams, F. (1993) "Fodor's Modal Argument." *Philosophical Psychology* 6 (1), pp.41-56.
- Alston, W. (2005) "Perception and Representation." *Philosophy and Phenomenological Research* LXX (2), pp.253-289.
- Aydede, M. (1997) "Has Fodor Really Changed His Mind on Narrow Content?" *Mind & Language* 12 (3/4), pp.422-458.
- Bach, K. (1982) "De re Belief and Methodological Solipsism." In Woodfield, A., ed. *Thought And Object: Essays on Intentionality*. Oxford: Clarendon Press, Oxford University Press, pp.121-151.
- Bach, K. (1987) *Thought and Reference*. Oxford: Clarendon Press.
- Bach, K. (1988) "Burge's New Thought Experiment: Back To The Drawing Room." *Journal of Philosophy* 85, pp.88-97.
- Baker, L.R. (1987a) "Content By Courtesy." *Journal of Philosophy* 84, pp.197-213.
- Baker, L.R. (1987b) *Saving Belief: A Critique of Physicalism*. Princeton: Princeton University Press.
- Barrett, J. (1997) "Individualism And The Cross Contexts Test." *Pacific Philosophical Quarterly* 78, pp.242-260.
- Biro, J. (1992) "In Defence Of Social Content." *Philosophical Studies* 67, pp.277-293.
- Blackburn, S. (1984) *Spreading The Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.
- Block, N. (1986) "Advertisement for a Semantics for Psychology." In French, P.A., T.E. Uehling, jr. and H.K. Wettstein, eds. *Midwest Studies in Philosophy Volume X: Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press, pp.615-678.
- Block, N. (1987) "Functional Role And Truth Conditions." *The Aristotelian Society - Supplementary Volume LXI*, pp.157-181.
- Block, N. (1990) "Inverted Earth." In Lycan, W.G., ed. *Mind And Cognition: An Anthology*. 2<sup>nd</sup> ed. Oxford; Malden, MA: Blackwell Publishers, pp.484-498.

- Block, N. (1991) "What Narrow Content is Not." *In* Loewer, B. and G. Rey, eds. *Meaning In Mind: Fodor and his Critics*. Oxford UK & Cambridge MA: Blackwell, pp.33-64.
- Braddon-Mitchell, D. and F. Jackson. (1996) *The Philosophy of Mind and Cognition*. Oxford; Cambridge, Massachusetts: Blackwell Publishers.
- Brown, C. (1986) "What Is a Belief State?" *In* French, P.A., T.E. Uehling, jr. and H.K. Wettstein, eds. *Midwest Studies in Philosophy Volume X: Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press, pp.357-378.
- Brown, C. (1993) "Belief States and Narrow Content." *Mind & Language* 8 (3), pp.343-367.
- Brueckner, A. (2001) "Defending Burge's Thought Experiment." *Erkenntnis* 55, pp.387-391.
- Brueckner, A. (2003) "Contents Just Aren't In The Head." *Erkenntnis* 58, pp.1-6.
- Burge, T. (1979) "Individualism and the Mental." *In* French, P.A., T.E. Uehling, jr. and H.K. Wettstein, eds. *Midwest Studies in Philosophy Volume IV: Studies in Metaphysics*. Minneapolis: University of Minnesota Press, pp.73-121.
- Burge, T. (1982) "Other Bodies." *In* Woodfield, A., ed. *Thought And Object: Essays on Intentionality*. Oxford: Clarendon Press, Oxford University Press, pp.97-120.
- Burge, T. (1986) "Individualism And Psychology." *Philosophical Review* 95 (1), pp.3-45.
- Burge, T. (1988a) "Cartesian Error and the Objectivity of Perception." *In* Grimm, R.H. and D.D. Merrill, eds. *Contents of Thought*. Tucson: The University Of Arizona Press, pp.62-76.
- Burge, T. (1988b) "Reply: Authoritative Self-Knowledge and Perceptual Individualism." *In* Grimm, R.H. and D.D. Merrill, eds. *Contents of Thought*. Tucson: The University Of Arizona Press, pp.86-98.
- Burge, T. (1989) "Individuation and Causation in Psychology." *Pacific Philosophical Quarterly* 70, pp.303-322.
- Burge, T. (1995) "Reply: Intentional Properties and Causation." *In* Macdonald, C. and G. Macdonald, eds. *Philosophy of Psychology: Debates on Psychological Explanation, Volume One*. Oxford UK & Cambridge USA: Blackwell, pp.226-235.
- Byrne, A. (2001) "Intentionalism Defended." *The Philosophical Review* 110 (2), pp.199-240.
- Campbell, J. (1982) "Extension And Psychic State: Twin Earth Revisited." *Philosophical Studies* 42, pp.67-89.

- Carnap, R. (1947) *Meaning And Necessity*. Chicago, IL: University of Chicago Press.
- Carter, R. (2000) *Mapping the Mind*. Phoenix.
- Chalmers, D.J. (2002) The Components of Content. Arizona: University of Arizona, Department of Philosophy. Available from: <http://www.u.arizona.edu/~chalmers/papers/content.html>. [Accessed on 05.07.2003]
- Chalmers, D.J. (2003a) "The Nature Of Narrow Content." *Philosophical Issues* 13, pp.46-66.
- Chalmers, D.J. (2003b) The Representational Character of Experience. Arizona: University of Arizona, Department of Philosophy. Available from: <http://www.u.arizona.edu/~chalmers/papers/representation.html>. [Accessed on 05.07.2003]
- Cheng, K-Y. (2002) "Narrow Content And Historical Accounts: Can Fodor Live Without Them?" *Journal of Philosophical Research* 27, pp.101-113.
- Comesana, J. (2005) "Justified vs. Warranted Perceptual Belief: Resisting Disjunctivism." *Philosophy and Phenomenological Research* LXXI (1), pp.367-383.
- Craig, E. ed. (1998) *Routledge Encyclopedia of Philosophy*. London; New York: Routledge.
- Crane, T. (1991) "All The Difference In The World." *Philosophical Quarterly* 41 (162), pp.1-25.
- Crane, T. (1998) "Intentionality." In Craig, E., ed. *Routledge Encyclopedia of Philosophy*. London; New York: Routledge, pp.816-821.
- Cummins, R. (1989) *Meaning and Mental Representation*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press.
- Cummins, R. (1991) "Methodological Reflections on Belief." In Bogdan, R., ed. *Mind And Common Sense*. Cambridge: Cambridge University Press, pp.53-70.
- Dancy, J., ed. (1988) *Perceptual Knowledge*. Oxford: Oxford University Press.
- Davidson, D. (1987) Knowing One's Own Mind." In Lycan, W.G., ed. *Mind And Cognition: An Anthology*. 2<sup>nd</sup> ed. Oxford; Malden, MA: Blackwell Publishers, pp.383-395.
- Davies, M. (1991a) "Individualism and Perceptual Content." *Mind* C (4), pp.461-484.
- Davies, M. (1991b) "Perceptual Content And Local Supervenience." *Proceedings Of The Aristotelian Society* LXXXII, pp.21-45.

- Davies, M. and I.L. Humberstone (1980) "Two Notions of Necessity." *Philosophical Studies* 38, pp.1-30.
- Dennett, D.C. (1981) "True Believers: The Intentional Strategy and Why It Works." In Rosenthal, D.M., ed. *The Nature of Mind*. New York; Oxford: Oxford University Press, pp.339-350.
- Dennett, D.C. (1982) "Beyond Belief." In Woodfield, A., ed. *Thought And Object: Essays on Intentionality*. Oxford: Clarendon Press, Oxford University Press, pp.1-95.
- Dennett, D.C. (1987) *The Intentional Stance*. London; Cambridge MA: A Bradford Book, The MIT Press.
- Devitt, M. (1981) *Designation*. New York: Columbia University Press.
- Devitt, M. (1984) "Thoughts and Their Ascription." In French, P.A., T.E. Uehling, jr. and H.K. Wettstein, eds. *Midwest Studies in Philosophy Volume IX: Causation and Causal Theories*. Minneapolis: University of Minnesota Press, pp.385-420.
- Devitt, M. (1990) "A Narrow Representational Theory of the Mind." In Lycan, W.G., ed. *Mind And Cognition: A Reader*. Oxford; Cambridge, MA: Basil Blackwell, pp.371-398.
- Dretske, F. (1995) *Naturalizing the Mind*. Cambridge, MA: The MIT Press.
- Edwards, S. (1994) *Externalism in the Philosophy of Mind*. Aldershot: Avebury.
- Egan, F. (1991) "Must Psychology Be Individualistic?" *Philosophical Review* 100 (2), pp.179-203.
- Egan, F. (1992) "Individualism, Computation, and Perceptual Content." *Mind* 101, 443-459.
- Egan, F. (1995) "Computation and Content." *Philosophical Review* 104 (2), pp.181-203.
- Evans, G. (1973) "The Causal Theory Of Names." In Moore, A.W., ed. *Meaning And Reference*. Oxford: Oxford University Press, pp.208-227.
- Evans, G. (1982) *The Varieties of Reference*. Oxford: Clarendon Press, Oxford University Press.
- Farkas, K. (2003) "What Is Externalism?" *Philosophical Studies* 112 (3), pp.187-208.
- Farkas, K. (2008) "Phenomenal Intentionality Without Compromise." *The Monist* 91 (2), pp.273-293.
- Field, H. (1977) "Logic, Meaning and Conceptual Role." *Journal of Philosophy* 74, pp.379-409.

- Field, H. (1978) "Mental Representation." *Erkenntnis* 13, pp.9-61.
- Fodor, J.A. (1975) *The Language of Thought*. Sussex: Harvester.
- Fodor, J.A. (1980) "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *In Representations: Philosophical Essays on the Foundations of Cognitive Science*. The Harvester Press, pp.225-253.
- Fodor, J.A. (1982) "Cognitive Science and the Twin-Earth Problem." *Notre Dame Journal of Formal Logic* 23 (2), pp.98-118.
- Fodor, J.A. (1984) "Semantics, Wisconsin Style." *In A Theory of Content and Other Essays*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press, pp.31-49.
- Fodor, J.A. (1985) "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum." *In A Theory of Content and Other Essays*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press, pp.3-29.
- Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press.
- Fodor, J.A. (1989) "Substitution Arguments and the Individuation of Beliefs." *In A Theory of Content and Other Essays*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press, pp.161-176.
- Fodor, J.A. (1990) *A Theory of Content and Other Essays*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press.
- Fodor, J.A. (1991a) "A Modal Argument for Narrow Content." *In Macdonald, C. and G. Macdonald, eds. Philosophy of Psychology: Debates on Psychological Explanation, Volume One*. Oxford UK & Cambridge USA: Blackwell, pp.206-225.
- Fodor, J.A. (1991b) "Replies." *In Loewer, B. and G. Rey, eds. Meaning In Mind: Fodor and his Critics*. Oxford UK & Cambridge MA: Blackwell.
- Fodor, J.A. (1994) *The Elm and the Expert*. Cambridge, MA: MIT Press.
- Fodor, J. and E. Lepore (1992) *Holism: A shopper's guide*. Oxford; Cambridge USA: Blackwell.
- Forbes, G. (1998) "Proper Names." *In Craig, E., ed. Routledge Encyclopedia of Philosophy*. London; New York: Routledge, pp.752-757.
- Frances, B. (1999) "On The Explanatory Deficiencies Of Linguistic Content." *Philosophical Studies* 93, pp.45-75.
- Frege, G. (1892) "On Sense And Reference." *In Moore, A.W., ed. Meaning And Reference*. Oxford: Oxford University Press, pp.23-42.

- Gallois, A. (1998) "Sense-Data." In Craig, E., ed. *Routledge Encyclopedia of Philosophy*. London; New York: Routledge, pp.694-698.
- Georgalis, N. (1999) "Rethinking Burge's Thought Experiment." *Synthese* 118, pp.145-164.
- Georgalis, N. (2003) "Burge's Thought Experiment: Still In Need Of Defense." *Erkenntnis* 58, pp.267-273.
- Gray, R. (2003) "Tye's Representationalism: Feeling The Heat?" *Philosophical Studies* 115, pp.245-256.
- Gray, R. (2005) "On The Concept Of A Sense." *Synthese* 147, pp.461-475.
- Gray, R. (2006) "Natural phenomenon terms." *Analysis* 66 (2), pp.141-148.
- Hannan, B. (1994) *Subjectivity & Reduction: An Introduction to the Mind-Body Problem*. Boulder; San Francisco; Oxford: Westview Press.
- Harman, G. (1982) "Conceptual Role Semantics." *Notre Dame Journal of Formal Logic* 23 (2), pp.242-256.
- Heil, J. (1992) *The Nature of True Minds*. Cambridge: Cambridge University Press.
- Horgan, T. and J. Tienson. (2002) "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In Chalmers, D., ed. *Philosophy Of Mind: Classical And Contemporary Readings*. Oxford: Oxford University Press, pp.520-533.
- Horgan, T., Tienson, J. and G. Graham. (2004) "Phenomenal Intentionality and the Brain in a Vat." In Schantz, R., ed. *The Externalist Challenge: New Studies on Cognition and Intentionality*. Berlin: de Gruyter, pp.297-317.
- Horowitz, A. (1996) "Putnam, Searle, And Externalism." *Philosophical Studies* 81, pp.27-69.
- Horowitz, A. (2001) "Contents Just Are In The Head." *Erkenntnis* 54, pp.321-344.
- Horowitz, A. (2005) "Externalism, The Environment, And Thought-Tokens." *Erkenntnis* 63, pp.133-138.
- Jackendoff, R. (1991) "The Problem of Reality." *Nous* 25, pp.411-433.
- Jackson, F. (1976) "The Existence Of Mental Objects." In Dancy, J., ed. *Perceptual Knowledge*. Oxford: Oxford University Press, pp.113-126.
- Jackson, F. (1998) "Reference And Description Revisited." *Philosophical Perspectives* 12, pp.201-218.



- Jackson, F. (2001) "Locke-ing onto Content." In Walsh, D.M., ed. *Naturalism, Evolution and Mind*. Cambridge: Cambridge University Press, pp.127-143.
- Jackson, F. (2003) "Representation And Narrow Belief." *Philosophical Issues* 13, pp.99-112.
- Jackson, F. and P. Pettit (1988) "Functionalism and Broad Content." *Mind* 97 (387), pp.381-400.
- Jackson, F. and P. Pettit (1993) "Some Content is Narrow." In Heil, J. and A. Mele, eds. *Mental Causation*. Oxford: Clarendon Press, pp.259-282.
- Jacob, P. (1990) "Externalism Revisited: Is There Such A Thing As Narrow Content?" *Philosophical Studies* 60, pp.143-176.
- Kaplan, D. (1989) "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In Almog, J., J. Perry and H. Wettstein, eds. *Themes From Kaplan*. New York; Oxford: Oxford University Press, pp.481-563.
- Katz, J.J. (1990) "The Domino Theory." *Philosophical Studies* 58, pp.3-39.
- Kitcher, P. (1985) "Narrow Taxonomy And Wide Functionalism." *Philosophy of Science* 52, pp.78-97.
- Kriegel, U. (2002) "Phenomenal Content." *Erkenntnis* 57, pp.175-198.
- Kriegel, U. (2007) "Intentional Inexistence and Phenomenal Intentionality." *Philosophical Perspectives* 21, pp.307-340.
- Kriegel, U. (2008) "The Dispensability of (Merely) Intentional Objects." *Philosophical Studies* 141, pp.79-95.
- Kriegel, U. and T. Horgan (2008) "The Phenomenal Intentionality Research Program." University of Arizona. Available from: [uriakriegel.com](http://uriakriegel.com).
- Kripke, S.A. (1971) "Identity And Necessity." In Moore, A.W., ed. *Meaning And Reference*. Oxford: Oxford University Press, pp.162-191.
- Kripke, S.A. (1979) "A Puzzle About Belief." In Salmon, N. and S. Soames, eds. *Propositions And Attitudes*. Oxford: Oxford University Press, pp.102-148.
- Kripke, S.A. (1980) *Naming and Necessity*. Oxford: Blackwell.
- Lepore, E. and B. Loewer (1986) "Solipsistic Semantics." In French, P.A., T.E. Uehling, jr. and H.K. Wettstein, eds. *Midwest Studies in Philosophy Volume X: Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press, pp.595-614.

- Levine, J. (2003) "Experience and Representation." In Smith, Q. and A. Jokic, eds. *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press, pp.57-76.
- Lewis, D. (1979) "Attitudes *De Dicto* and *De Se*." In *Philosophical Papers: Volume I*. New York; Oxford: Oxford University press, pp.133-156.
- Lewis, D. (1980) "Veridical Hallucination And Prosthetic Vision." In Dancy, J., ed. *Perceptual Knowledge*. Oxford: Oxford University Press, pp.79-91.
- Lewis, D. (1994) "Reduction of mind." In Lewis, D. *Papers in metaphysics and epistemology*. Cambridge: Cambridge University Press, pp.291-324.
- Loar, B. (1982) "Conceptual Role and Truth-Conditions." *Notre Dame Journal of Formal Logic* 23 (3), pp.272-283.
- Loar, B. (1987) "Subjective Intentionality." *Philosophical Topics* 15 (1), pp.89-122.
- Loar, B. (1988a) "Social Content and Psychological Content." In Grimm, R.H. and D.D. Merrill, eds. *Contents of Thought*. Tucson: The University Of Arizona Press, pp.99-110.
- Loar, B. (1988b) "Reply: A New Kind of Content." In Grimm, R.H. and D.D. Merrill, eds. *Contents of Thought*. Tucson: The University Of Arizona Press, pp.121-139.
- Loewer, B. (1982) "The Role of "Conceptual Role Semantics"." *Notre Dame Journal of Formal Logic* 23 (3), pp.305-315.
- Loewer, B. and G. Rey, eds. (1991) *Meaning In Mind: Fodor and his Critics*. Oxford UK & Cambridge MA: Blackwell.
- Lycan, W.G., ed. (1990) *Mind And Cognition: A Reader*. Oxford; Cambridge, MA: Basil Blackwell.
- Lycan, W.G. (1996) *Consciousness and Experience*. Cambridge, Mass.: MIT Press.
- Lycan, W.G., ed. (1999) *Mind And Cognition: An Anthology*. 2<sup>nd</sup> ed. Oxford; Malden, MA: Blackwell Publishers.
- Lycan, W.G. (2000) *Philosophy Of Language: A contemporary introduction*. London and New York: Routledge.
- Macdonald, C. (1995) "Introduction: Anti-individualism and Psychological Explanation." In Macdonald, C. and G. Macdonald, eds. *Philosophy of Psychology: Debates on Psychological Explanation, Volume One*. Oxford UK & Cambridge USA: Blackwell, pp.156-172.
- Macdonald, C. (1998) "Externalism and Norms." In O'Hear, A., ed. *Current Issues in Philosophy of Mind*. Cambridge: Cambridge University Press, pp.273-301.

- Macpherson, F. (2005) "Colour Inversion Problems for Representationalism." *Philosophy and Phenomenological Research* LXX (1), pp.127-152.
- Maloney, J.C. (1991) "Saving Psychological Solipsism." *Philosophical Studies* 61, pp.267-283.
- Manfredi, P.A. (1993) "Two routes to narrow content: both dead ends." *Philosophical Psychology* 6 (1), pp.3-21.
- Margolis, E. and S. Laurence (2002) "Lewis' Strawman." *Philosophical Quarterly* 52 (206), pp.55-65.
- Matthews, R.J. (1988) "Comments on Burge." In Grimm, R.H. and D.D. Merrill, eds. *Contents of Thought*. Tucson: The University Of Arizona Press, pp.77-86.
- McCulloch, G. (1995) *The Mind and Its World*. London and New York: Routledge.
- McDermott, M. (1986) "Narrow Content." *Australasian Journal of Philosophy* 64 (3), pp.277-288.
- McDowell, J. (1977) "On The Sense And Reference Of A Proper Name." In Moore, A.W., ed. *Meaning And Reference*. Oxford: Oxford University Press, pp.111-136.
- McDowell, J. (1982) "Criteria, Defeasibility, And Knowledge." In Dancy, J., ed. *Perceptual Knowledge*. Oxford: Oxford University Press, pp.209-219.
- McDowell, J. (1984) "De Re Senses." In *Meaning, Knowledge, And Reality*. Cambridge, MA; London: Harvard University Press, pp.214-227.
- McDowell, J. (1986) "Singular Thought And The Extent Of Inner Space." In Pettit, P. and J. McDowell, eds. *Subject, Thought, and Context*. Oxford: Clarendon Press, pp.137-168.
- McDowell, J. (1994) *Mind and World*. Cambridge: Harvard University Press.
- McGinn, C. (1982) "The Structure of Content." In Woodfield, A., ed. *Thought And Object: Essays on Intentionality*. Oxford: Clarendon Press, Oxford University Press, pp.207-258.
- McGinn, C. (1989) *Mental Content*. Oxford; New York: Basil Blackwell.
- McKinsey, M. (1991) "The Internal Basis Of Meaning." *Pacific Philosophical Quarterly* 72, pp.143-169.
- McKinsey, M. (1993) "Curing Folk Psychology Of 'Arthritis'." *Philosophical Studies* 70, pp.323-336.
- McKinsey, M. (1994) "Individuating Beliefs." *Philosophical Perspectives* 8, pp.303-330.

- Mill, J.S. (1843) *A System of Logic, Ratiocinative and Inductive*. London: Longmans, Green & Co., published 1884.
- Millikan, R.G. (1984) *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R.G. (1993) *White Queen Psychology, and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Moore, A.W., ed. (1993) *Meaning And Reference*. Oxford: Oxford University Press.
- Nagel, T. (1974) "What is it like to be a bat?" *Philosophical Review* 83, pp.435-450.
- Noonan, H.W. (1986) "Russellian thoughts and methodological solipsism." In Butterfield, J., ed. *Language, Mind and Logic*. Cambridge: Cambridge University Press, pp.67-90.
- Noonan, H.W. (1991) "Object-Dependent Thoughts And Psychological Redundancy." *Analysis* 51 (1), pp.1-9.
- Noonan, H.W. (1993) "Object-Dependent Thoughts: A Case of Superficial Necessity but Deep Contingency?" In Heil, J. and A. Mele, eds. *Mental Causation*. Oxford: Clarendon Press, pp.283-308.
- Patterson, S. (1990) "The Explanatory Role Of Belief Ascriptions." *Philosophical Studies* 59, pp.313-332.
- Pereboom, D. (1995) "Conceptual Structure And The Individuation Of Content." *Philosophical Perspectives* 9, pp.401-428.
- Perlis, D. (1991) "Putting One's Foot in One's Head – Part I: Why." *Nous* 25, pp.435-455.
- Perry, J. (1977) "Frege On Demonstratives." In Yourgrau, P. ed. *Demonstratives*. Oxford: Oxford University Press, pp.50-70.
- Perry, J. (1979) "The Problem Of The Essential Indexical." In Salmon, N. and S. Soames, eds. *Propositions And Attitudes*. Oxford: Oxford University Press, pp.83-101.
- Perry, J. (1998) "Broadening the Mind." *Philosophy and Phenomenological Research* LVIII (1), pp.223-231.
- Pettit, P. and J. McDowell, eds. (1986) *Subject, Thought, and Context*. Oxford: Clarendon Press.
- Putnam, H. (1973) "Meaning And Reference." In Moore, A.W., ed. *Meaning And Reference*. Oxford: Oxford University Press, pp.150-161.

- Putnam, H. (1975) "The meaning of 'meaning'." In Putnam, H. *Mind, Language and Reality: Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press, pp.215-271.
- Putnam, H. (1981) *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Putnam, H. (1983) "Computational Psychology and Interpretation Theory." In Rosenthal, D.M., ed. *The Nature of Mind*. New York; Oxford: Oxford University Press, pp.527-535.
- Putnam, H. (1988) *Representation and Reality*. Cambridge, Massachusetts; London: A Bradford Book, The MIT Press.
- Putnam, H. (1999) *The Threefold Cord: mind, body, and world*. New York: Columbia University Press.
- Quine, W.V. (1956) "Quantifiers and the Propositional Attitudes." In Rosenthal, D.M., ed. *The Nature of Mind*. New York; Oxford: Oxford University Press, pp.306-311.
- Recanati, F. (1990) "Direct Reference, Meaning, and Thought." *Nous* 24, pp.697-722.
- Recanati, F. (1994) "How Narrow is Narrow Content?" *Dialectica* 48 (3-4), pp.209-229.
- Rechenauer, M. (1997) "Individualism, Individuation And That-Clauses." *Erkenntnis* 46 (1), pp.49-67.
- Rey, G. (1992) "Semantic Externalism And Conceptual Competence." *The Aristotelian Society* LXXXII, pp.315-331.
- Rey, G. (1998) "A Narrow Representationalist Account Of Qualitative Experience." *Philosophical Perspectives* 12, pp.435-457.
- Rosenthal, D.M., ed. (1991) *The Nature of Mind*. New York; Oxford: Oxford University Press.
- Russell, B. (1905) "On Denoting." In *Logic and Knowledge: Essays 1901-1950*. London: George Allen & Unwin Ltd., pp.41-56.
- Russell, B. (1910/11) "Knowledge By Acquaintance And Knowledge By Description." In Salmon, N. and S. Soames, eds. *Propositions And Attitudes*. Oxford: Oxford University Press, pp.16-32.
- Salmon, N. and S. Soames, eds. (1988) *Propositions And Attitudes*. Oxford: Oxford University Press.

- Searle, J.R. (1983) *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Segal, G. (1989a) "The Return of the Individual." *Mind* 98 (389), pp.39-57.
- Segal, G. (1989b) "Seeing What Is Not There." *Philosophical Review* 98 (2), pp.189-214.
- Segal, G. (1991) "Defence of a Reasonable Individualism." *Mind* C (4), pp.485-494.
- Segal, G. M. A. (2000) *A Slim Book about Narrow Content*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press.
- Senor, T.D. (1992) "Two-factor theories, meaning holism, and intentionalistic psychology: a reply to Fodor." *Philosophical Psychology* 5 (2), pp.133-151.
- Siewert, C. (1998) *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.
- Silverberg, A. (1994) "Meaning Holism And Intentional Content." *Pacific Philosophical Quarterly* 75, pp.29-53.
- Silverberg, A. (1995) "Narrow Content: A Defense." *The Southern Journal of Philosophy* 33, pp.109-127.
- Smith, Q. and A. Jokic, eds. (2003) *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- Sosa, E. (1970) "Propositional Attitudes *De Dicto* and *De Re*." *Journal of Philosophy* LXVII.
- Sosa, E. (1993) "Abilities, Concepts, and Externalism." In Heil, J. and A. Mele, eds. *Mental Causation*. Oxford: Clarendon Press, pp.309-328.
- Stalnaker, R.C. (1989) "On What's in the Head." In *Context And Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press, pp.169-193.
- Stalnaker, R.C. (1990) "Narrow Content." In *Context And Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press, pp.194-209.
- Stalnaker, R. (1991) "How to Do Semantics for the Language of Thought." In Loewer, B. and G. Rey, eds. *Meaning In Mind: Fodor and his Critics*. Oxford UK & Cambridge MA: Blackwell, pp.229-237.
- Stalnaker, R.C. (1993) "Twin Earth Revisited." In *Context And Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press, pp.210-221.
- Stalnaker, R.C. (1999) *Context And Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

- Stich, S.P. (1978) "Autonomous Psychology and the Belief-Desire Thesis." In Lycan, W.G., ed. *Mind And Cognition: A Reader*. Oxford; Cambridge, MA: Basil Blackwell, pp.345-361.
- Stich, S.P. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, Massachusetts; London: A Bradford Book/The MIT Press.
- Stich, S.P. (1986) "Are Belief Predicates Systematically Ambiguous?" In Bogdan, R.J., ed. *Belief*. Oxford: Clarendon Press, pp.119-147.
- Stich, S.P. (1991) "Narrow Content meets Fat Syntax." In Loewer, B. and G. Rey, eds. *Meaning In Mind: Fodor and his Critics*. Oxford UK & Cambridge MA: Blackwell, pp.239-254.
- Strawson, P.F. (1979) "Perception And Its Objects." In Dancy, J., ed. *Perceptual Knowledge*. Oxford: Oxford University Press, pp.92-112.
- Stroll, A. (1998) *Sketches of Landscapes: Philosophy by Example*. Cambridge, MA; London: A Bradford Book, The MIT Press.
- Taylor, K. (1989a) "Narrow Content Functionalism and the Mind-Body Problem." *Nous* 23, pp.355-372.
- Taylor, K. (1989b) "Supervenience And Levels Of Meaning." *The Southern Journal of Philosophy* 27 (3), pp.443-458.
- Taylor, K.A. (1998) "Propositional Attitude Statements." In Craig, E., ed. *Routledge Encyclopedia of Philosophy*. London; New York: Routledge, pp.771-779.
- Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA; London: A Bradford Book, The MIT Press.
- Tye, M. (1998) "Inverted Earth, Swampman, And Representationism." *Philosophical Perspectives* 12, pp.459-477.
- Tye, M. (2000) *Consciousness, Color, and Content*. Cambridge, MA: The MIT Press.
- Velmans, M. (1998) "physical, psychological and virtual realities." In Wood, J., ed. *The virtual embodied: presence/practice/technology*. London and New York: Routledge, pp.45-60.
- Wakefield, J.C. (2002) "Broad versus narrow content in the explanation of action: Fodor on Frege cases." *Philosophical Psychology* 15 (2), pp.119-133.
- Wallace, J. and H.E. Mason (1990) "On Some Thought Experiments about Mind and Meaning." In Anderson, C. and J. Owens, eds. *Propositional Attitudes*. CSLI, pp.175-199.

White, S.L. (1982) "Partial Character And The Language Of Thought." *Pacific Philosophical Quarterly* 63, pp.347-365.

White, S.L. (1992) *The Unity Of The Self*. MIT Press.

Wiggins, D. (1990) "Putnam's Doctrine Of Natural Kind Words And Frege's Doctrines Of Sense, Reference, And Extension: Can They Cohere?" *In* Moore, A.W., ed. *Meaning And Reference*. Oxford: Oxford University Press, pp.192-207.

Woodfield, A. (1982) "On Specifying the Contents of Thoughts." *In* Woodfield, A., ed. *Thought And Object: Essays on Intentionality*. Oxford: Clarendon Press, Oxford University Press, pp.259-297.

Yourgrau, P. ed. (1990) *Demonstratives*. Oxford: Oxford University Press.

Zemach, E.M. (1976) "Putnam's Theory On The Reference Of Substance Terms." *Journal of Philosophy* 73, pp.116-127.