



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Journal for the Theory of Social Behaviour

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa40167>

Paper:

Amadae, S. (2018). Game theory, cheap talk and post-truth politics: David Lewis vs. John Searle on reasons for truth-telling. *Journal for the Theory of Social Behaviour*
<http://dx.doi.org/10.1111/jtsb.12169>

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.

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

**Game Theory, Cheap Talk and Post-Truth Politics:
David Lewis vs. John Searle on Reasons for Truth-Telling**

“The fabric of civilian life is now wrapped in a linguistic fog of war.”

—John Herrman, March 14, 2017, *NYT*

I offer two potential diagnoses of the behavioral norms governing post-truth politics by comparing the view of language, communication, and truth-telling put forward by David Lewis (extended by game theorists), and John Searle. My first goal is to specify the different ways in which Lewis, and game theorists more generally, in contrast to Searle (in the company of Paul Grice and Jürgen Habermas), go about explaining the normativity of truthfulness within a linguistic community. The main difference is that for Lewis and game theorists, “truthful” signaling follows from an alignment of interests, and deception follows from mixed motives leading to the calculation that sending false information is better for oneself. Following in the Enlightenment tradition, Searle argues that practical reasoning, which involves mastery of at least one language, requires that actors intend to communicate. This intention includes constraining the content of statements to uphold veracity conditions.

After distinguishing between these two accounts, I will articulate the implications for explaining, and even informing actions, constitutive of post-truth politics. I argue that the strategic view of communication is sufficient neither to model everyday conversation nor to reflect a public sphere useful for democratic government. In game theory most communication is “cheap talk,” and the delivery of either “true” or “false” statements that can be tested against world states is simply a matter of maximizing expected utility. For Searle, communication presupposes the intention to utter meaningful statements that satisfy truth conditions against world states. While animals can engage in deceptive acts by signaling misleading information, humans lie in the sense that they intentionally try to pass-off a *false* but *meaningful* statement as a true statement which is an ostensibly contradictory endeavor. Strategic manipulation of information is only possible by piggybacking mendacious intention upon linguistic conventions which presume truthful use of terms to make communication possible. As

communication increasingly is mediated by electronic information technologies, that automate social relations by transforming the immediacy of I-You interactions into digitized protocols, the game theoretic understanding of prevails. Conversing is converted into signaling and meaning reduces to information. Agents, who may be anonymous or artificial, offer no warrant of truthfulness. Conveying meaning becomes a means to satisfying strategic ends, such as profitability, that disrupt the inherent telos of communication.

I. Motivation: “Post-Truth Politics”

Since the United Kingdom Brexit Vote, and the US presidential election of Donald Trump there has been a proliferation of public instances of and reactions to demonstrably false statements of fact. There are both critical analyses of the phenomenon and queries regarding how we can explain its emergence and by now wide-spread existence. Some blame the failure of neoliberal economic policies, and the European and North American public’s new disdain for the validity of economic science and its policy advice. This could explain populism and the Trump administration’s appeal to it by brandishing post-truth rhetoric that “vilifies elites, [and] combined with glorification of the people...[, expresses] a pretty strong anti-intellectualism and skepticism about expertise” (Steven K. Bannon quoted by Marc Fisher in the *Washington Post*, “The political lexicon of a billionaire populist,” March 9, 2017). Expressive demonstration of support for Trump, then takes the form of acceding to even blatantly obvious untruths, including the size of the inaugural crowd gathered on the Washington Mall on Jan. 20th 2017 corroborated by graphic photographs (Fandos 2017).

The commentary on the “Post-Truth Era” began prior to 2016, as is evident by the 2004 book bearing this title. This author, Ralph Keyes reports that since the turn to this century, and probably with gradually increasing frequency over the decades prior, lying has become normalized. There are similar accounts of an escalation in cheating (Callahan 2006). This consideration leads me to consider the advent of a post-truth era of politics from a wider perspective that poses the question of what cultural anchors we have that would uphold the norm of truth-telling in public and private life? This shifts attention from the empirical investigation of what the contemporary phenomenon is,

how it is expressed, how frequent it is, and how it can be explained to posing a different question. I ask what theoretical tools exist within the mainstream academy to ground a concept of truth and a theory of why an individual would be and should be truthful in communication. This is not a new topic and harkens back to the Enlightenment division of labor between scientific impartiality supporting public matters of fact on the one hand (Shapin and Shaffer 1985) and of the predominant coincidental virtue theory of ethics on the other that deemed lying, and falsely stating intentions by breaking promises, breaches of ethical imperatives (Driver 2003). Although it is not possible to prove that changes in ideational superstructure transform how actors interpret the meanings of their actions and also potentially their behavior (see McCloskey 2016 for an attempt), it is at least possible to ask what intellectual theory concurs with a practice.

This paper investigates what culture of truth-telling rational choice theory, also referred to as game theory, decision-theory, and strategic rationality, supports if its axioms of choice informed rather than described human action. My tentative hypothesis is that it offers theoretical support which could be used to rationalize the post-truth individual and collective action we currently observe. At a minimum, we expect a breakdown of communication to occur in polarized situations of fundamental conflict. At a maximum we could normalize a view of strategic rationality that accepts that every utterance must contain accurate, false, or ambiguous information dependent on the reward structure of interactions all referred to as games. Countering this, Searle, Grice, and Habermas argue that communication presupposes the intention to convey meaning that in turn relies on speakers' commitment to truthfulness. Their analysis is necessary to mobilize a definition of truth-telling as opposed to deceptive behavior and its analytic integrity is superior to that of game theory's reduction of communication to cheap talk.

II. Communication as signaling and cheap talk

I first articulate a simple, intuitively accessible analysis of how the game theoretic account of communication as signaling works. This can be inferred from David Lewis' early treatment of language and communication using game theory and developing the concept of *Convention*, the name of his book (1969, 122-202). As I do

below, Lewis divides the labor of explaining linguistic communication into two halves, the first discussing individuals' decisions to tell the truth or be deceptive in particular cases of communication, and the second providing a game theoretic account for the existence of language as a widespread social institution. Throughout Lewis views a communicative act as one of transmitting a signal that, assuming that actors develop a common basis to associate meaning with a signal, can be communicated truthfully in accord with the conventional meaning, or falsely at odds with the linguistic convention. Thus every communicative act is one of signaling. Signals fulfill an instrumental function for the sender who seeks to instantiate a belief about the world in another actor with the aim of getting that actor to do that which is most beneficial for the sender. Every use of communication in game theory is instrumental. Acts of communication are either "cheap," making no difference to the payoffs defining games, or costly, directly impacting an actor's gain. Thus, an example of cheap talk is saying, "I love you." A costly act of signaling is bombing an adversary as a signal that worse will come if they do not comply with one's demands. Here the signal, in the form of bombing a city, leads to a new outcome that directly impacts actors' expected utility, and so is not considered "cheap talk" in the parlance of game theory. An act of sacrifice, such as fasting, to demonstrate loyalty is similarly not cheap talk because the act exacts a price; it is a costly signal. However, most communication is classified as cheap talk in game theory, which "consists of costless, nonbinding, nonverifiable messages that may affect the listener's beliefs" (Farrell and Rabin 1996, 116).

The picture of the world is straightforward, with three basic assumptions underlying it. First, the signaler perceives a state of the world; this perception counts as a piece of information about a world state. Second, the signaler associates a symbol with this world state. If the signaler is the member of a common linguistic community, he knows the symbol that conventionally obtains to convey that state of the world. Third, the signaler who is in an interaction, referred to as a game, with another agent, knows the possible outcomes of the game, and how the outcome of the interaction could be manipulated depending on what the receiver believes about the world as a result of a signal. Signaling games, as these situations are referred to, assume an asymmetry of information such that the signaler knows something about the world,

which could be a concrete fact, or a subjective state of intending to take a certain action. The signaler has knowledge which could be useful for the receiver of the information. Deciding how to communicate in terms of which signal to send, and whether it should be accurate or deceptive, is a matter of strategic calculation. Another type of asymmetry of information pertains to whether every player knows what branch, or sequence of actions, has occurred in order to make the best strategic choice; in another case, knowing what one actor will do could make it obvious for the other to choose her action.

For simplicity, I first introduce a coordination game, in which both agents individually do best when the other also does best. In this example, one agent uses a signal to convey whether the state of the world has a population of one or two of an entity in question, such as enemies. The signaler can signal either “one,” or “two.” The sender and the receiver know the conventional meaning of “one” and “two,” and moreover know that the payoff for both the sender and the receiver is positive if they converge on the meaning of “one,” or “two,” but negative if they are mismatched, and the signaler sends “one,” and the receiver understands “two.” In Lewis’ *Convention* (1969) receiving a signal, that is relying on it to determine what to believe about what state of the world obtains, translates directly into acting (1969, 124-125). In this game (Table 1), if I accept the signal “one” to mean “one” attacker, then I act to coordinate with the sender to neutralize a single opponent. If I accept the signal “two” to mean “two attackers,” then I coordinate with the sender to counter one of two opponents.

[Insert Table 1 Here]

This case shows that given this reward structure, both actors benefit by converging to take actions that respect a common linguistic convention.

Game theorists divide the world into situations in which all actors’ interests align, referred to as coordination games, situations in which actors’ interests are diametrically opposed, called conflict games, and mixed motive games with elements of conflict and coordination. Only in some game settings will communication, assumed to be cheap, make a difference. If communication makes no difference in the outcome,

then it is not only cheap, but also inconsequential. An example of this type is the Prisoner's Dilemma game, used to model many situations throughout human affairs, from international relations security dilemmas to the social contract, public goods, and collective action, to noncooperative bargaining and even marriage. The payoff structure, or individuals' expected gain, depending on what outcome obtains, has the following characteristic form. If both actors cooperate, they can achieve an outcome of joint remuneration. If both actors defect, they lose this opportunity and leave with what value they had. If one actor cooperates, but the other defects, then the cooperator loses capital which the defector walks away with without expending any effort. The following game payoff matrix (Table 2) obtains.

[Insert Table 2 Here]

Given the much studied dissatisfying outcome, or Nash mutual-best-reply equilibrium, of the Prisoner's Dilemma game, standard game theory texts discuss whether the two agents could be better off if only they had communicated, signaling their intention to cooperate. Were the agents to cooperate in this example, then both Sarah and Ralph would leave the encounter having gained 50 utility units each, while if they both defect, each receives 25. Rational players will have the motive to send deceptive information that they intend to cooperate, hoping that the other actor may likewise cooperate. This would expose the other player to being suckered, because the signaler's true intent, consistent with her preferences, is to defect no matter what. However, in this well formulated interaction defined by agents' comprehensive, all inclusive, preferences over outcomes, it is not credible to tell the other one will cooperate: the pay offs contain all the information about what motivates actors. Saying that one will cooperate contradicts one's expected utility. Thus even if both signal the intent to cooperate, both will defect, and the outcome will be 25 utils for each (see also Farrell and Rabin, 1996, 112-113). This is quite a startling result given that bargains between two actors over any set of goods typically have the Prisoner's Dilemma structure from within the game theory paradigm (Nemeth 1972). The usual remedy for this within game theory is either to introduce an external sanctioning body that threatens each actor with a penalty if they defect, or to assume the two actors meet each

other repeatedly and for an indefinitely long length of time (Axelrod 1984). Although the latter does not guarantee that they will manage to achieve cooperation some or most of the time, it at least introduces that possibility. Still communication is superfluous to this result.

It is worth introducing one more example from the game theoretic cheap talk literature to make clear how pay offs structure encounters, and the content of communication is strictly calculated from the role that it could play in achieving the best possible outcome for the individuals who signal. In the role of receiver, each has to calculate whether the transmitted information is credible, revealing true information about the world, or lacks credibility. This likelihood of credibility is wholly a function of the reward structure of the game. In some signaling games, or situations in which transmitting information could affect what actions individuals take and hence what outcome will obtain, sending information in accordance with a probabilistic function for truth or falsehood can also be beneficial to the signaler. In this case the way to determine the content of a message relies on making decisions according to probabilities and leads to messages in which its substance conveys no meaning.

In this next example, an individual is better off by deceiving the other actor about the true state of the world. A job candidate and an employer have the following payoff function over the game structured by the company's assessment of whether the candidate has high-level or low level-skills. It is beneficial for the candidate to misrepresent her skills as high because she will receive a higher paying job wherein she can get by, with her low skill-set remaining undiscovered. This game matrix (Table 3, Farrell and Rabin 1996, 106) is as follows.

[Insert Table 3 Here]

The game reward structure depends on whether the candidate actually has high or low skills, and whether the employer offers the candidate the high level or the low level position. Here the candidate benefits from the employer believing her skill-set is high, regardless of the truth. If, in fact, the candidate has less skill, the employer gains less from this situation because poor performance for higher compensation results. In the

case that the candidate has high-level skills, then it is not good for her or the company if the employer believes she has low skills.

The game theoretic analysis explores many more possibilities, and identifies cases in which communication as cheap talk makes a difference to the outcomes and when it does not. The more conflict, the less difference signaling makes. The full range of situations includes cheap talk about private information; cheap talk as babbling (speakers utter nonsense or random statements); cheap talk about actors' intentions; and cheap talk in situations in which coordinating on a shared plan is good for both, but one plan is better for one individual, and the second plan better for the second. Cheap talk helps the actors converge to mutually optimal outcomes in the case that messages are self-signaling and self-committing. The first means that signaling and acting are for all practical purposes equivalent; the second means that were the receiver to believe the signal sent and act accordingly, then it is automatically in the sender's best interest to do what she said she would do.

However, one needs to read carefully through the game theoretic analysis to understand when it pays to talk using shared linguistic conventions, when it may be as good to babble, and when it is useful to send either a true or false signal. Were we interested in learning to communicate more effectively, and read game theory texts accordingly as tutorials, we learn that:

These three examples [varying degrees of alignment between a signaler's interest in revealing the truth versus the receiver's obvious interest in knowing the truth] suggest some general principles. Sometimes there is no incentive to lie, and cheap talk will fully convey private information. If there is too strong an incentive to lie, cheap talk becomes meaningless. However, even if there is some limited incentive to lie, cheap talk can convey some meaning in equilibrium. (Farrell and Rabin 1996, 107)

In a strategic context, such as game theory assumes best characterizes all encounters, bold overt lying that is obviously in an actor's interest may not pay off because it lacks credibility, and could, one assumes, lead the signal receiver to be skeptical even were the statement to be accurate. This situation may require a non-cheap talk remedy: speaking with actions that alter the agent's costs and benefits, including, for example, that a candidate enrolls in an expensive educational program to achieve credentials so that her skill set and self-reporting match. Surprisingly, communication can play a role

in the employer and employee example discussed above (Table 3) to achieve a stable, “equilibrium,” outcome in the case that cheap talk is permitted to be imprecise. In this case the candidate can make ambiguous statements about her abilities just sufficient to make the employer give her the benefit of the doubt. In this game theoretic analysis, if the degree of exaggeration is precise, then the employer can calculate in reverse exactly how much to disregard the exaggerated skill claims to offset them.

Game theorists draw conclusions about the role of communication in games, reinforcing that the more actors’ interests are aligned, the greater the constructive role of communication can be in their achieving mutually beneficial outcomes. However, in situations where actors’ interests are not perfectly aligned, there can be many contexts in which signaling inaccurate information about states of the world or intentions can lead to better results for the signaler. Throughout the game theoretic treatment of communication as signaling and cheap talk, the standard assumption is often made that the pay offs for outcomes are well-defined for each player, and common knowledge among them. Furthermore, only a finite set of signals can be sent.

It is usually standard to assume that if actors have access to what game theorists refer to as a “rich language,” such as English, in which although they are able to send deceptive information, they are not able to send an incomprehensible signal (Farrell and Rabin 1996, 110). Signalers can send words that can be understood as nonsensical messages. This set of considerations could be important in contexts in which remaining silent is the best option for the sender who will get the benefit of the doubt if it is impossible to determine which world state of which only she knows, obtains. The rich language assumption, without which many games would have no ground for modeling cheap talk, is useful in games wherein actors take pure strategies without randomizing action choices. It also supports babbling equilibrium in games solved using mixed strategies. In a mixed strategy-equilibrium, which must be assumed throughout game theory to guarantee the existence of solutions to games, each player chooses which action to take based on a randomizing function. One example is, “play strategy *A* 30% of the time, and play strategy *B* 70% of the time.” Often at least one or several mixed, hence randomized, strategy sets among players provides a game solution (Nash equilibrium), although these may well be suboptimal. In the case wherein an actor will

choose an action at random, there is no way for the accompanying signal to convey either accurate or inaccurate information about which act will obtain. Thus, if actors randomize signals in mixed-strategy equilibrium, their signals are referred to as “babble” because they are independent from what state will obtain, and convey no information (Heath 2001, 70). Being formed of perfect grammatical strings the signals are comprehensible, yet they lack credibility (Farrell and Rubin, 1996, 110). This occurs if I use one die role to determine what I will do, and a different die role to determine what I will say I will do.

From this brief overview of how game theory treats communication, it is possible to grasp the need to postulate a clear reward structure for all encounters modeled as games, and to treat the use of communication as dependent on the incentives that motivate actors. All communication is cheap, unless it directly alters the potential outcomes. In some contexts, such as those involving conflict, signals about world states will not alter the outcome of the interaction which instead will follow entirely from pay offs. In other contexts, specifically those in which actors’ interests are wholly or significantly aligned, signals may be self-signaling (read off from or directly indicated by an action choice) and self-committing (because truthfully inducing a credible belief about a world state in another actor gives the signaler a reason to commit to the act she signaled she would take). (Stalnaker 2006, 93)

If we divide our population into the signaler (S) and the receiver (R) and conceive of a signaling game in which S can be of one of several types that refer to which act she will take, then we can define rational signals and credible signals (Stalnaker 2006, 92-93).

A message is *prima facie rational* (pf rational) for player *S*, of type *t* if and only if *S* prefers that *R* believe the content of the message.

Hence the definition of *prima facie rational* counts as rational any signal that the sender wishes the receiver to believe, although a rational message could have true content, or false content. Furthermore:

A message is *credible* if and only if it is pf rational for some types, and only for types for which it is true.

Hence, for a message to be credible, it has to be rational for the sender to send it. Here the crucial part of the definition is that the message is known to be credible, or not solely, primarily from the perspective of the sender who knows her type, and knows if the message is true with respect to either her intentions to act or a fact about the world. The implication is that the receiver can only deduce a signal's credibility from a deductive analysis of the game's payoff structure.

There is a third point, that adds the consideration that credible messages are those that the receiver will believe:

It is common belief that the content of any credible message that is sent is believed by R.

This means that it is a common basis for belief among actors that it is rational for the receiver of a credible message to believe it. The problem with this latter point is that, since credibility is asymmetrically known to the sender, it may not be clear that the receiver will have sufficient information to determine which messages are credible. However, in the cases of deduced credibility, the message will be believed by the receiver and acted on accordingly. Messages are designated as credible, incredible (demonstrably false) or ambiguous wholly depending on the payoff structure of the game. Within game theory there is no association of credible signaling with, for example, an actor's characteristic qualities or virtues. Deciding to reveal the truth given common linguistic references, remaining silent, babbling, or sending false information is strictly a strategic calculation.

Furthermore, despite the exacting formalism of game theory, and despite setting up signaling games so that communication plays the role of treating language use as instrumental action that solely functions to maximize agents' expected utility, it remains underspecified what messages should be sent, and whether the receiver should find them credible (Farrell and Rabin, 1996). In game theory, there is only a finite set of possible messages that can be sent. Thus, whereas the sender can determine what possible messages are credible insofar as they are both true and if believed will lead the receiver to take an action that furthers the sender's interests, still there will be signaling

games in which “*R* may be mistaken or uncertain about whether a message is credible, [and] *S* may be unsure whether a credible message will in fact be believed” (Stalnaker 2006, 96). Messages must be deciphered with respect to what actors’ interests are, and what the sender stands to gain by inducing belief in the mind of a receiver.

III. Language as an Institution

Game theorists have tended to follow the philosopher of language Paul Grice (1989) in distinguishing between individual signaling acts and the existence of any language convention that a particular signaling act relies on to associate a symbol with a state of affairs in the world. Although not using game theory, Grice treated speech acts as fundamentally instrumental: “A speech act is an action that like any rational action should be explained in terms of the purposes for which it is performed [namely conveying meaning to instill a belief about a world state in the receiver’s mind], and the agent’s beliefs about its consequences” (Stalnaker 85). Importantly, Grice argued that “speech is an institution whose function is to provide resources to mean things, and that what it is to mean things needs to be explained independently of the institution whose aim it is to provide the means to do it” (Stalnaker 2006, 85). Leaving aside the personification of language as though it has the function to enable actors to convey meanings, I have followed this division of labor between analyzing individual acts of communication and considering language as an institution, by first discussing how individuals convey meanings in signaling games. Signaling games can be of various types, and all communication, classified as cheap talk, must be one of these types. This requires that “the meaningfulness of the speech act [is] dependent on the payoff structure of the game” (Heath, 2001, 70). There is no non-instrumental communication. All speech acts are a function of agents’ assessment of which signal best serves their interests given the need to at a minimum send a plausibly credible message or no message. Communication is a subset of rational action that is thoroughly strategic.

Next I consider what resources game theory has to explain the institution of language. For the reason that game theory analyzes micromotives to explain interactions of any size, any game theoretic account of language necessarily must factor

in how communication works in varying interactions. David Lewis offered the first account of language using game theory, and it heeds us to notice some simplifications that he made to make it feasible to develop this reductionist account (Heath 2001, 79) of speakers' meaningful utterances in communicative actions. Many of the assumptions are already specified above: that the pay offs are well specified for each actor as part of individuals' complete and consistent evaluation of the worth of all possible outcomes; players have common knowledge of each others' pay offs; and actors are rational insofar as they act to maximize expected utility. In order to ensure that games have solutions, it is also necessary to allow that actors can play mixed strategies by introducing randomized decision-procedures. As well, agents in orthodox game theory only value outcomes and not the means by which they are achieved (Hausman, 2012, 53). Due to the existence of babbling equilibria, or the need to resort to randomized statements as cheap talk when actors play randomized strategies, Lewis ignored mixed strategy games. This is because these games' solutions seem to render linguistic communication nonsensical, Lewis ruled them out of consideration. This theoretical move was strictly one of convenience and shows how limiting a theory of language reduced to strategic rationality is: game theory cannot guarantee solutions (equilibria) without the device of mixed-strategies (Heath 2001, 70). Hence a theory of language relying on game theory that disregards this foundational solution using mixed strategies may be of limited value. Lewis and others moved ahead with providing a strategic account of language by focusing on coordination games. Thus his book *Convention* only studies coordination games, and seeks to explain the basis of language as a convention derived from humans' need to coordinate actions using common symbols of reference to send signals.

Furthermore following Lewis, literature explaining language in terms of strategic rationality treats it as a convention emerging from coordination problems. This approach rests on several points. Most importantly and most obviously, following Lewis, in any conventional game all actors' preferences over outcomes align so that they have a joint interest in coordinating their actions. In this case, to use the vocabulary developed above, signals are self-signaling and self-committing. Consider a choice of signal A and signal B, and a conventional interpretation that A means up and B means down. In a coordination game, when the sender signals either A or B, it is rational for

the receiver to deduce that the signal is credible, and to act accordingly. Joseph Heath uses the example of a builder who needs blocks and slabs, and calls out to an assistant “block” when a block is required, and “slab” when a slab is required, and the assistant duly delivers the block or the slab in response. However, of course while two communicators can converge on a meaning, it is unclear how a single linguistic convention obtains among an entire population. Heath shows how even a slight change in the builder example, that introduces cross-purpose motives, brings communication based on this model to a standstill. Let us suppose that the builder learns that he is being paid by the hour rather than by the job, and the assistant’s payment remains by the job. Now the builder has the motive to stretch out the time for the project, and sometimes has the incentive to send the wrong signal about whether he needs a slab or a block, leading to inefficient time use. As Heath points out, “simply by acquiring this new incentive, all his speech suddenly becomes meaningless” (2001, 71). Messages have strings of characters that convey a meaning in accordance to the conventional structure of a rich background language; yet they break with the convention to send deceptive information. Hence the overall expression becomes meaningless. Heath’s point is that while the meaningfulness of communicative acts relies on the existence of conventions, without aligned motives to converge on the use of terms, communication rapidly breaks down.

One possible explanation for language to exist as a convention follows if the majority of interactions are coordination interactions because “linguistic conventions can exist only where there is a coordination problem” (Heath 2001, 71). As soon as we allow that there are many types of interactions throughout society, this calls into questions that there exists a sufficient critical mass of coordination situations to ensure that linguistic conventions can be maintained. Given diverse types of interactions, confusion over terms will arise from either cheap talk that is incredible, or meaningless, as in the babbling equilibrium that coincides with games solved through mixed strategies, or simply deceptive.

Lewis, who is an astute and thorough philosopher, does address these concerns because he acknowledges that, given the normal assumption that Prisoner’s Dilemmas may abound throughout social interactions (1969, 91), that it is possible that most

individuals prefer others to state the truth about states of affairs, while they themselves may prefer to lie (1969, 182). He attempts to derive a predilection to be truthful from individuals' overall preference to live in a society with meaningful communication over a state without linguistic conventions. Without this overall preference to live in a society with the mutual expectation of being truthful, communication could likely breakdown. Given the large degree of interactions routinely modeled as mixed motive games, it is not surprising that, in the end, Lewis looks to agents to be "habitually truthful" in a linguistic community, and "accustomed to expect truthfulness...on the part of others" (183). In setting up his argument with these default assumptions, he begs the question of how the institution of language can arise and be maintained by simply observing that it is everyone's rational interest that others tell the truth. Thus he obviates the detailed analysis of most signaling games (eg. Farrell and Rabin 1996), and hence reveals the deeper stakes of accepting the game theoretic account of communication. If individuals either do not habitually, or by obligation, tell the truth, then people's common linguistic convention for communicating may deteriorate into nonsense.

The key point is that, as opposed to John Searle's account of both language as an institution and individual speech acts, according to Lewis the strategic account of signaling encompasses both language as a conventional institution *and* agents' intentions underlying their engaging in signaling. Lewis states:

Searle draws this moral from his example...: "we must capture both the intentional and the conventional aspects [of communication] and especially of the relationship between them."

Lewis goes on to conclude, "I have been arguing that once we capture the conventional aspect we are done. We have captured the intentional aspect as well" (159). The intentional aspect of language refers to how a sender deliberately sends a signal, whether true or false and credible or incredible, and the receiver reacts to it by believing it or not and modifying his action or not, *because he recognizes that the sender intended to communicate* (see Lewis 152-159). In game theory, this means that the receiver realizes that the signal comes from the sender who intends to alter his belief state, and interprets it thus. Here Lewis explicates Paul Grice who originally defines a special case of

meaning something in his statement, “We may say that ‘A meant_{NN} something x’ is roughly equivalent to ‘A uttered x’ with the intention of inducing a belief by means of the recognition of his intention” (Grice 1989, 219). In game theory, the intentional aspect of communication lies in aiming to influence the receiver’s beliefs to better realize the sender’s goals as specified in the sender’s payoffs.

Acquiescing to the game theoretic account of language, which Lewis specifies for us encompasses the intentional aspect of communication, accepts an instrumental view in which all rational action is motivated by the beliefs and desires of the actors. Lewis and game theorists aim “to explain the distinctive character of *communicative* action, taking for granted the normal resources for the explanation of rational action—beliefs, desires, values, and ends, intentions” (Stalnaker 2006, 86). However, as Lewis argued in pioneering this approach, the conventional approach to language built up from coordination games captures, without extra consideration, the intentional aspect.

One half of communicating is signaling a message that is designed to affect the belief state of a fellow interactant to better satisfy the signaler’s preferences. The other half involves that the receiver must recognize the signal as an intention to communicate and decide whether to accept it as credible information and accordingly alter his beliefs and actions. Within game theory, then, the signal must be recognized by the receiver as deriving from the signaler’s expected utility from sending a message insofar as it is aimed to alter the receiver’s beliefs about world states. However, again, the sender’s intention is indistinguishable from examining her preferences over the game’s outcomes. Intentionality will have a wholly different role in Searle’s theory of language because communication is impossible without presupposing the concepts of meaning and truth. Thus where the game theoretic account of communication as signaling derives the intentional aspect of language from actors’ strategic satisfaction of desires given their beliefs, we will see that by contrast, for Searle desires and beliefs are a form of intentional state that reflect propositional content. Formulating and evaluating propositional statements such as “it is raining,” or “I will take an umbrella” have relationships to the world and require commitment to be accepted as true.

Before moving onto Searle, let’s consider, for the sake of argument, the lessons we would learn about communication if we accept the veracity of the game theoretic

account. This involves accepting the overall theory of agency proposed by strategic rationality and applying it to communicative acts. Agents have complete and consistent preference rankings of outcomes over all possible world states (Lewis 1969). Agents either function in a world with a common background language, or without. Every signal sent with the intention that the interactant partner will receive, understand, and believe the signal, is consistent with the consideration of maximizing expected utility. Sending a true or false signal, or choosing to send no signal or a babbling signal, are a function of the payoffs to games. Signaling games have asymmetric information, so the sender knows which signals are credible and which are not. The receiver needs to study the game structure, its pay off matrix, to discern whether signals are credible, plausibly credible, and may offer any useful information for making a choice. We are encouraged to learn more about signaling games in order to interpret signals:

The framed degree in your doctor's office, the celebrity endorsement of a popular cosmetic, and the telephone message from an old friend are all signals. The signals are potentially valuable because they allow you to infer useful information. These signals are indirect and require interpretation. They may be subject to manipulation.... (Sobel 2008)

The author concludes that, "The theory of signaling games is a useful way to describe the essential features of all three examples," and we may assume is also useful for deciding how best to send and interpret signals in order to realize goals, (Sobel 2008). Thus we are naïve to function in our contemporary world without studying signaling games that assess actors' propensity to conform to a norm of truthfulness based on the pay offs to interactions.

We can regard the game theorists' objective of studying communications charitably via several observations. I doubt that most theorists investigating linguistic communication entertain the possibility that theoretical accounts could transform language use. Rather game theorists attempt to systematically model communication as it exists. Second, having been applied to human rationality and to the behavioral regularities of animals in evolutionary contexts, game theory offers a naturalistic means of understanding communication. This is in line with the recognition that few believe that language was intentionally created, rather it must have emerged from individual

interactions over millennia stretching back into prehistory. Third, given that strategic rationality offers at least a plausible account of rational action and communication, it is worth investigating the insights that it produces.

IV. John Searle's Intentionality, Action and Meaning

Maintaining consistency with scientific naturalism, Searle puts forward his position on rationality to challenge what he refers to as “the Classical Model.” He counters Donald Davidson and Bernard Williams’s approaches to human rationality that require that all motives be part of agents’ motivational set of desires. Whereas Searle’s identification of the Classical Model includes Davidson’s causal decision theory, and Williams’s internalist ascription of reasons underlying action, game theory offers even a more limited account of rationality. Whereas Davidson requires that agents’ master language and make propositional statements before they formulate preferences in terms of whether they prefer or disprefer propositional statements to be true, in strategic rationality, actors’ preference rankings over world states are primitive (Heath 2001 19-23; Davidson 1984, 1990). No symbolic mediation between the perceptual recognition of a world state and ranking of its value against other world states is required. The symbolic representation of world states is either directly caused through habitual neurocognition (Gershman et al. 2015), or by a convention which is similarly experienced as a habitual causal process. This, then, permits Lewis and game theorists to build communication and language up from games which are stipulated by elementary pay off matrices. Furthermore, while Searle seeks to refute Williams’ argument, that the only motivating features of action are internal reasons deriving from agents’ desires. Thus he introduces desire-independent reasons, the types of desires that orthodox game theory recognizes are more restricted. This is a subtle point about the nature of expected utility theory necessary to solve games: it demands that actors’ only have preferences over path-independent outcomes and lotteries of path-independent outcomes (Heath 2001; Guala 2006, 258; Hausman 2011). This is important because decision theory broadly construed, which could accommodate Davidson and likely Williams as well, permits, for example, that if an agent prefers to keep a promise, this preference can be reflected in the agent’s utility function (Hubin 2001). However, in

game theory, promising-keeping cannot be reflected in expected utility functions. This follows because the utterance of promising can only be cheap talk, unless it is accompanied by a costly signal, such as opening an escrow account to hold funds that will transfer when goods are delivered. Making a promise alters nothing about the outcomes over which actors' have rankings. Thus, for example, in the Prisoner's Dilemma situation with tangible rewards over which the actor has the appropriate utility function, telling the other person "I promise to cooperate" alters nothing about the reward structure, and can play no motivational role in altering the agent's behavior. Thus, game theory offers a more reductionist account of action than Searle's broader target of the Classical Model of rationality.

Searle introduces the concept of "intentional states," which are aspects of mental states that "are directed at, or about, or of states of affairs in the world beyond themselves" (2001, 34). Intentional states include, but are not limited to, intending. They can have propositional form, and these have a structure analogous to speech acts. Beliefs and desires are intentional states and their content in the form of propositions. Propositions state the content of belief, and are not the object of belief. There are "two directions of fit," using Searle's terminology, between the world and the mind. Beliefs, specifically true beliefs, are predicated on mental states accurately reflecting the state of the world that they are about. Here Searle introduces the claim that "a belief is a commitment to truth" insofar as it involves accepting, or recognizing, a proposition. Already to formulate a belief requires a propositional form dependent on access to a rich background language that has stipulated meanings for terms. Thus in accepting a belief the individual must assess if it satisfies conditions of veracity in order to commit to its being true. Alternatively stated, in order to distinguish between beliefs and true beliefs requires an individual's commitment to ascertaining that the proposition reflecting the intentional state of belief satisfies validity conditions.

For the reason that theorists do not doubt that desires motivate action, and that the Classical Model, or even stripped down game theoretic strategic rationality model, captures this aspect of desire satisfaction, Searle is more concerned to defend a theory of communication that is not instrumental. He goes further to defend that actors can have desire-independent reasons for actions. In this latter respect, he permits that, beyond

what game theory accepts, actors can have a broad complex of desires that encompass considerations outside merely those consistent with orthodox expected utility theory. Thus, he accepts a broad view of “all encompassing desires” as do Davidson and Williamson, yet he denies that this is sufficient to account for motivations or reasons for action such as obligations to tell the truth or to keep promises (169-171). Searle holds that even a reason for believing a true proposition is desire independent because agents may prefer to be in a state of denial or delusion.

Desires have the opposite “direction of fit” than true beliefs because rather than accurately reflecting world states, they represent how actors would like world states to be (37). Thus beliefs are true or false, and desires are either fulfilled or unfulfilled. Whereas the game theoretic model of action tends to assume that beliefs could come about as a representational mental state caused by external stimuli (Gershman et al. 2015), and views desire plus belief as a reason that causes action, Searle’s understanding of action is more complex. He argues that there are three “gaps” between causal physical processes and mental states that are crucial to how we understand and motivate action (49-51). One is between the perceptual phenomena that are the basis on which we formulate our beliefs, and our acceptance of true beliefs as propositions that satisfy validity claims. The second is in deciding what desires to fulfill. Although this second consideration may sound trivial, even in decision theory it is not trivial to decide between two equally preferred outcomes, leaving aside long term and short term considerations, risk and uncertainty, and different orders of considerations from the sensual to the other-regarding. Third, even after having decided on a course of action, still the agent needs to initiate and maintain the sequence of actions to achieve her goal, regardless of unexpected occurrences or challenges. These gaps reflect that forming propositions to satisfy truth conditions diverges from a causal theory of perception and natural meaning in which a sign, such a predator’s footprint, initiates a causal chain such as a fight-or-flight instinct.

Searle’s theory of rationality and action contains more topics than is necessary to discuss here because my focus is on communication and truth-telling. Viewing this topic from the pedagogic perspective I am adopting here, I concentrate only on what is necessary for actors to understand a robust alternative perspective from that given by

strategic rationality. Thus I am leaving aside Searle's account of subjective and objective epistemological and ontological points of view. The key concept for Searle with respect to mobilizing his account of communication is meaning. On this topic it becomes possible to directly compare Searle's argument with that of Lewis and game theorists. We recall Grice's idea that communication must be mutually recognized and must convey the intention of the speaker, aka signaler in game theory. Again, in Grice's terms:

We say that 'A meant_{NN} [non-natural meaning] something by x' is roughly equivalent to 'A uttered x' with the intention of inducing a belief by means of the recognition of this intention. (Grice 1989, 219).

Although as we shall see ahead, Searle and game theorists give alternative rationales for how one should decide whether to communicate a truth or falsehood, it seems that philosophers of language do agree on one point. This is that regardless of whether the content of a communicated message is true or false, the sender cannot deceive the receiver about the meaning of the message. In other words, even though communications can be "devious and deceptive," a speaker "cannot attempt to deceive her interlocutor about what she intends him to understand her to be meaning" (Stalnaker 2006, 86). Thus the underlying instrumental function of language is to convey meaning.

We have seen how Lewis builds up meaningful communication from actors' convergence on associating a symbolic reference to a particular world state across different members of a population who interact in games. If games are pure coordination situations in which actors' preferences are perfectly aligned, then it will be in everyone's interest to converge to the same reference between symbolic representations and specific world states. However, I previously discussed how it is well-accepted that as soon as there are situations of pure conflict or partial conflict, not only will there be lack of convergence, but actors may be motivated to send false signals and will need to calculate which signals are credible.

Mobilizing a theory of language and communication from a different point of departure, Searle stipulates that, "meaning is a matter of the intentional imposition of

conditions of satisfaction on conditions of satisfaction” (53). This conceptualization of meaning needs to be explained. It has two stages, one is that a speaker makes a statement conveying meaning, and hence intends to communicate. This is in keeping with Grice’s instrumental understanding of communication. However, making an utterance, and making an utterance with meaning are two different considerations. Because Searle assumes the propositional content of beliefs which in turn presupposes a rich background language with common terms of expression, conveying a meaning requires satisfying the conditions of world-fit grounding the veracity of the propositional statement. If the agent (a) intends to make an utterance; (b) seeks to communicate meaning to a listener; and (c) seeks to deceive, or induce a false belief, in the listener, then the strategic purpose of the speaker is parasitic on both the institution of language as a cooperative basis for expressing meaning and on the speaker’s intention to convey meaning. Another way of saying this is that using language to convey false propositional claims in order to attempt to induce a false belief in a listener provides an instance of failing to satisfy the veracity conditions of a proposition, which is a precondition for the establishment of meaning in the first place. This is not to state that lying is not possible. But it does provide us with the conditions to distinguish between lying and deception, which can occur in nature through biological signaling (Zollman et al. 2012). This deception is different from intentionally misusing language as a system available to convey meaning by simulating the effect of communicating meaning, but actually undermining it.

Whether or not one accepts Searle’s approach to rationality, language, and communication is not significant here. Rather my point is to contrast a pedagogy of the types of reasons that can be marshaled to explain to individuals why they should tell the truth. Obviously, a tutorial in cheap talk and game theory offers a categorically distinct rationale from that offered by Searle. Searle, who already argues that true belief involves commitment (for the reason that false beliefs can be comfortable, and even self-promoting), and that lying is an abuse of communication which has the point of conveying meaning, goes further to argue that individuals have an obligation to tell the truth, just as they have an obligation to abide by promises made. Searle’s argument for the latter builds on the case he has already constructed, discussed above. Here Searle

moves in the opposite direction to Lewis. The latter attempted to build up an explanation for language based on linguistic conventions, or convergence to fixed meanings for symbols in reference to world states, through coordination games. Searle adopts a different tack by building on his idea that true beliefs entail commitment and that intentionality requires recognition.

The idea of recognition seems new, but was already introduced by Grice who holds that for communication to occur both the speaker and listener must recognize that intentional meaning is conveyed from the first individual to the second. Here Searle goes beyond the idea that intending to convey meaning implicitly requires, in order to count as communicating, using terms in conventionally understood ways consistent with individuals' background mastery of language. He moves in the opposite direction to Lewis because rather than recognition merely being already contained within the intentional preference satisfaction of conveying a message to induce a belief-state in a listener, for Searle the intention to convey meaning signifies realizing what conditions of veracity for true propositions must be met. This is what is required to articulate an utterance that conveys meaning. Searle argues that simple deception, or sending a false signal in line with strategic rationality to induce a false belief state, hoping to trigger favorable behavior in a competitor, cannot count as intending to communicate because communicating within a linguistic community requires accepting the universality of conditions for language use. Lying presupposes the ability to convey meaning using a shared common language and recognition of the meaning of terms in that language.

Lewis' account may be attractive because it assumes so little: specifically there is no need for speakers to have lofty theories of linguistic use or to participate in maintaining the "salient norms and rules of communication" that facilitate conveying meaning (Balliet 2010, 48). Searle seems to presuppose that people must intend to communicate in order to make meaningful utterances and to interpret them. By contrast, Lewis' and game theorists' account of language as signaling, with only the desire to maximize expected utility underlying all communicative acts, cannot account for how language and communication exist due to mixed motive games and randomized strategy equilibria. Communication among animals is treated on par with communication among humans (e.g. Zollman et al. 2012).

V. Taking a Position: Is the Strategic or Meaning-Based Account of Communication more Coherent?

How would one decide whether the strategic or meaning-based account of human communication is correct? To an extent, this question may be an empirical question about perception, cognition, neuroscience, and how humans actually conduct conversational exchanges. Given the enormity and dynamic quality of these overlapping fields of study, it would be difficult to take an empirical stand on this research as it relates on this question in this article. Thus I consider the question of the validity of the strategic versus meaning-based account of human communication from the perspective of the analytic coherence of the two approaches offered by game theory and John Searle complemented by Paul Grice. This section accepts Grice's instrumental approach to language and communication.

According to Grice, in everyday communication, as opposed to formal language, context is necessary to deduce the speaker's intended meaning. This is much like in game theoretic communication as cheap talk, listeners must deduce actors' intended meanings from the context: a game's payoff matrix. However for Grice, a necessary condition of everyday conversational exchange is speakers' warrant of the truthfulness of their statements due to their commitment to the cooperative principle of communication. Thus for Grice, conveying meaning depends on a commitment to tell the truth, among other linguistic norms. In game theory, by contrast, there are only linguistic conventions. Whether actors choose to conform to, or violate, them is strictly a matter of convenience. However, since conveying meaning depends on maintaining conventions, the game theoretic approach of sending signals and interpreting them based on incentives undermines any means to define truthfulness as upholding common standards of language use. Thus whereas for Grice speakers' credibility is a condition of communication, in game theory speakers' credibility is a condition of strategic calculation.

In reconstructing the game theoretic position on communication and language as fundamentally strategic, we can recognize that this approach takes a step beyond Grice. Grice put forward at least two important notions that Lewis built on to develop his game theoretic analysis of communication as convention. First, according to Grice the

meaning and truth-value of a sentence is distinct from *conversational implicature*, by which he refers to the interpreted meaning of sentences uttered and received within a specific conversational context. Grice holds that formal languages, necessary for definitively stipulating conditions that must hold for a proposition to be recognized as “true,” are fully interpretable. This means that the meaning of a sentence is intersubjectively beyond dispute, and that therefore the satisfiability of truth conditions is objective. However as we know, the conveyance of meaning in everyday conversation is much more complex to understand and also to theorize about. Here Grice’s second contribution is to argue that a speaker’s intention in making an utterance is essential to and yet wholly contained in that utterance. This permits the distinction between action A offering a sign, like a beheaded body, versus action B making a linguistic statement such as “he is dead.” Whereas the former leaves no ambiguity, with the latter the listener’s recognition of the speaker’s intent to communicate is crucial. In the former case meaning is natural and conveys information about a world state like, “those spots mean (meant) measles” has the structure “*x meant that p and x means p entail p*” (Grice 1957, 377). The difference between the first A case and the second B case is that in the A case recognition of the intention by the hearer is irrelevant to the success of the communicative act. The meaning that “he is dead” is naturally, subject to the laws of nature, entailed by *x*: in this case his being beheaded.

Game theory can work within this theoretical structure built by Grice, but it depends on an argumentative step that I suggest ultimately undermines Grice’s instrumental theory of language use. In signaling games, Grice’s concept of the importance of intention, insofar as the listener must realize she is receiving an intended meaning from the speaker, is upheld. This means that for the speaker, the intention to communicate is the aim of inducing a state of belief in a listener in order that the receiver uses this information to act in such a way to best achieve the sender’s aim. Likewise, the receiver knows that the sender’s signal is chosen with the aim of inducing a belief about the state of the world as a basis for the receiver’s action to satisfy the sender’s preferences.

Game theory goes beyond Grice because it requires lumping together determining the concordance or opposition of an uttered statement with linguistic

conventions, and the context of its utterance. In signaling games, an accurate, misleading or vague signal is selected by the sender and deciphered by the receiver wholly based on the payoff matrix, or “context” of the communication. According to Grice, in a formal language, conducive to scientific inquiry, the meaning of terms is fully interpretable and there are objective conditions that must be satisfied for propositions to be true. The context of conversation is superfluous because the truth and meaning of a statement are objective. Formal language use assumes either symmetry of access to information, or that a condition of epistemic virtue is upheld (see also Driver 2003). However, most linguistic exchange is not formal, and its context is important for listeners to understand the meaning communicated by speakers.

To make the distinction between Grice and game theory clear, consider the following case in which a speaker knows it is raining, and decides whether to communicate to an officemate the statement (a) “It is raining,” or (b) “It is not raining.” Within the context of a formal language this linguistic meaning code be coded to read (a₁) “P” or (b₁) “not-P.” Grice does not consider that formal languages useful for scientific inquiry could be used deceptively.¹ Thus in order to consider cheap talk we must turn to his account of conversational implicature, or communication in which the context is essential for the listener to deduce the meaning which the speaker attempts to convey (1975). In human communication, which game theory catalogues as cheap talk, context is necessary for the receiver in a signaling game to calculate whether the sender has the incentive to signal accurately, deceptively, or vaguely.

Consider the following “umbrella” signaling game in which the speaker (S) decides whether to communicate statements (a) or (b) above. The speaker’s goal is only to instrumentally communicate (a) or (b), but also with the overarching goal of inducing the listener to leave behind her umbrella on a day on which it is actually raining, so that S can take it for himself. Here the payoff matrix refers to the outcome in the actual case in which it is either raining or not raining, and the listener takes the appropriate action:

¹ In the context of formal systems with propositions reflecting world states, false statements are problematic because truth cannot be a property of a proposition, but only of whether a belief conforms to an existing world state. A false statement of fact exists as an object, but it is contradictory to assert a fact, which by definition has the existential quality of being, but of course, “simply has no being at all, if it is false” (Moore 1957, 262).

taking the umbrella when it rains, leaving it behind when it does not rain. From table 4 we can evaluate that (without considering future interactions) the payoff matrix is clear in instructing the speaker to make the misleading statement (b) on a rainy day on which taking the colleague's umbrella offers him the highest payoff.

[Insert Table 4 Here]

If the colleague is astute and realizes that she is participating in a signaling game, she will have to deduce this deception by perceiving that on a rainless day her co-worker gains nothing by communicating the fact that it is not raining, but stands to gain in the case that she is misled into believing it is not raining while it is actually raining. Thus, although game theorists assume that a natural language exists with sufficient specificity to understand the meaning of the two sentences "It is raining," and "It is not raining," the speaker's actual intended meaning to induce a false belief about the world in the listener can only be deduced from the specific strategic interaction given by the payoff matrix.

Grice defines that "meaning something by uttering x " holds if for audience A the utterer U makes statement x intending the following:

- (1) A to produce a particular response r
- (2) A to think (recognize) that U intends (1)
- (3) A to fulfill (1) on the basis of his fulfillment of (2). (1969, 151)

In his *Convention* Lewis seems to fulfill Grice's elementary definition of "meaning something" in conversation because utterers produce signals based on this structure of considerations, and listeners base their responses on these considerations as well.

However, as was made clear earlier, Lewis only considers coordination games in which actors' aims are aligned, so the umbrella signaling game is not a scenario he discusses.

In a coordination signaling game it is the case that when the audience understands A 's intended meaning and the response it is structured to achieve, A has an interest in supplying the intended reaction. However, in a mixed motive strategic game which involve a large percent of cases, this alignment of interests does not hold (see Zollman et al. 2012). This is because in these situations in which it behooves the speaker to be deceptive, meeting the conditions of (2) and (1) above only lead to (3) if the audience goes against its interests and acts in accordance with a false belief.

Thus it is a fair question whether, like Lewis, Grice only considers situations of mutually aligned interests. In reading Grice it is quickly clear that with respect to non-

formal languages equivalent to rich natural languages and everyday conversational exchanges, he stipulates that dialogue partners must uphold the *Cooperative Principle* (CP) (1975, 45). Grice agrees that communication is instrumental. A speaker's primary goal is to communicate the meaning of an utterance to a listener which in turn evokes a response. Grice presumes that this entails making truthful statements in view of the two *supermaxims* he introduces, (1) "do not say what you believe to be false"; and (2) "do not say that for which you lack adequate evidence" (46). He observes that in everyday conversation, "I expect your contributions to be genuine and not spurious" (45).

Whereas Lewis and Grice can concur in situations of coordination, as soon as there are mixed motives characteristic of many strategic games, Grice's framework for conversational implicature, or understanding the meaning of an utterance from its context, is irrelevant. This because Grice assumes that individuals in conversation have a "quasi-contractual" standing with each other similar to Searle's demand that speakers offer a warrant of truthfulness out of respect for their interlocutors and Habermas's argument that dialogue signifies participating in an I-you relationship that is categorically distinct from instrumentally using others primarily as a means to realize one's ends. Thus the strategic approach to communication, although finding common ground in identifying rationality with instrumental action, violates the first supermaxim not to make untruthful statements. Although admitting that many of us may maintain the Cooperative Principle out of habits learned in childhood, Grice regards the "standard type of conversational practice [consistent with the CP] not merely as something that all or most do IN FACT follow but as something that it is REASONABLE for us to follow, that we SHOULD NOT abandon" (sic 48). Here Grice's reference to being reasonable entailing the cooperative principle with commitment to truthfulness in communication resembles Searle and Habermas's position on rationality, as well as that of John Rawls (1985). For Searle, Habermas and Grice our actions not only can, but should, be motivated by consideration of reasonableness that encompass following the rules of norm-based practices that actors participate in. Within this normative framework for conversation, interlocutors can be called out for disregarding the overriding conversational norm of avoiding making false claims.

Thus we see that from an analytic perspective, game theory breaks with Grice's instrumental theory of communication because its followers have no means to challenge deceptive signaling by appealing to mutual recognition; their only recourse is to punish deceptive communicators. Grice goes further in positing that the natural default of participating in natural language exchange is to uphold the Cooperative Principle, and that abandoning the supermaxims is an act of opting out of everyday communication.

He concludes that:

I would like to be able to show that observance of the CP and maxims is reasonable (rational) along the following lines: that any one who cares about the goals that are central to conversation/communication (e.g., giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstances, in participation in talk exchanges that will be profitable only on the assumption that they are conducted in general accordance with the CP and maxims. (49)

Thus, as Habermas and Searle also argue, since the purposive act of communicating requires imparting interpretable meaning, the perverse case of using dialogue to convey misleading information can only be an ulterior goal that relies on the basic need to transmit meaningful statements. Linguists have difficulty categorizing a proposition uttered as a known falsehood.² As Habermas asserts, any deceptive use of language can only function effectively assuming that the underlying telos of communication itself to transfer meaning (Habermas 1984, 287).

VI. Conclusion

From the backdrop of the encompassing perspective argued by Habermas, the development of communicative rationality led to the self-organization of politics via a public sphere of discourse informed by independent newspapers and forums for forming public opinion leading to effective electoral politics. However, now non-linguistic exchange media threaten to colonize the "lifeworld" of communicative interactions of which the public sphere is one such domain. Habermas was specifically referring to money as the primary form of non-linguistic exchange media.

Whereas it is possible that early modern and modern markets may have exhibited coordination on prices in part due to communication (which game theory

² Since formal linguistic theorists assume that agents have an inherent interest in making true statements, the idea of purposefully stating known falsehoods as truths leads to the puzzle of the existential status of "false facts" (Moore, 1957, 261-266).

would refer to as cheap talk), not only large scale markets but also social media more comprehensively now rely on transmitted signals with only a remainder of cultural significance. They are largely disembodied of dialogic I-you relations fundamental for Grice, Searle, and Habermas. If we accept that there was an effective public sphere underlying the rise of liberal states from 1600 to 2000 that shared its basis with the epistemic virtue of truthfulness necessary for the institution of science and the normative rule of law (Habermas 1989), then it may be the case that the contemporary decay of the public into polarized campus incapable of constructive exchange could be explained by the new phenomenon of “automating the social.” Where Habermas challenges non-linguistic exchange media, with the automation of social networks, language itself is reduced to mere signaling and cheap talk. Processes for facilitating this transformation driving in the opposite direction to Habermas’ *Structural Transformation of the Public Sphere* are not only commodifying relationships but also treating language itself as a repository of symbols with connected meanings only useful for pursuing goals. Against the Enlightenment’s ambivalent generation of a free markets and democratic government, that were thought to reflect the end of history in 1989, we now are stuck with witnessing “reason’s inability to escape the slide into skepticism and doubt-that leads inevitably to moral chaos and the threat of illiberalism.” (Brown 2017 <http://iascculture.org/THR/channels/THR/2017/01/>)

Google Scholar demonstrates the widespread use of signaling games to model numerous institutions throughout political economy. In contemporary political practice actors have decided to promote false news as a means to realize their ends. False news can attract attention for being salacious, but it can also appear credible because of the irrationality of sending boldly false signals. If one believes the overall rationality of signalers, then their sending what appear to be patently incredible messages may be interpreted as instances in which these actors must have credible asymmetric knowledge. Furthermore, an actor’s interest may be in reaching an audience demographic who inadvertently signal their profile type by following false news stories. One prominent US corporation, 20th Century Fox, intentionally created fake news sources, and false news stories by providing advertising content to those pages because that way they reached their target demographic in 2016. Even though this advertising

campaign paid off financially, the company was moved to apologize with the observation that officials had not recognized the adverse implications of intentionally supporting the dissemination of false information, in this case the belief that vaccines are causally linked to autism, for the wider set of communicative practices that maintain civil society. Along with David Lewis, whose strategic account of individuals' signaling and the institution of language is inadequate, and who hoped that citizens would simply prefer to tell the truth in order to maintain a political community with a meaningful basis for communication, Paramount Pictures acknowledged that it must uphold a commitment to veracity in engaging the public in order to not contribute to the general corrosion of reducing linguistic conventions to post-truth politics (Maheshwari 2017).³

The mere fact that communication is increasingly computer mediated in venues in which social norms and rules are lax and virtually impossible to enforce by community members (Balliet 2010, 48-49) makes this topic more pressing now. Social media including Facebook and Twitter are concerned with their market share and profitability and thus sustain forums that embrace the broadcast of known falsehoods (*Economist*, 2017). In section V, I discussed how Grice and Searle support Habermas's conclusion that effective communication presupposes commitment to truthfulness that can help sustain a public sphere basic to civil politics (Habermas 1981, 1987, 1989). The binding force of communication lies in agents' ability to convey meaning, which assumes their commitment to warrant assertions with truthfulness and simultaneously offers respect to listeners as interlocutors, and not merely means to achieve the speakers' ends. Evidence shows that the implicit normative rules governing face-to-face discussion are "more likely to activate the social norm of promise keeping," which is a form of truthfully asserting that I will perform a future action (Balliet 2010, 48).

³ "Then came 20th Century Fox's [latest movie promotion tactic](#), for which it created fake news sites with plausible names like The Houston Leader and The Salt Lake City Guardian and populated them with partisan headlines involving Lady Gaga, President Trump, vaccinations and mental health. Many of the articles on the sites, which were apparently produced in collaboration with a fake news creator the company declined to identify, were shared by people on Facebook who apparently did not know they were part of a stunt meant to publicize the film 'A Cure for Wellness.'"

References

- Axelrod, Robert. 1984. *Evolution of Cooperation*. New York, Basic Books.
- Benz, Anton and Gerhard Jäger, Robert van Rooij. 2006. *Game Theory and Pragmatics*. ISBN: 978-1-349-52317-7 (Print) 978-0-230-28589-7 (Online)
- Callahan, David. 2004. *The Cheating Culture: Why More Americans are Doing Wrong to Get Ahead*. New York: Mariner Books.
- Davidson, Donald. 1995. "Could there be a Science of Rationality." *International Journal of Philosophical Studies*, 3:1, 1-16.
- Davidson, Donald. 1990. "The Folly of Trying to Define Truth." *Journal of Philosophy*, 93, 263-278.
- Davidson, Donald, 1984. "Truth and Meaning." In *Inquiries into Truth and Interpretation*. Oxford: Clarendon, 17-36.
- Davies, William, 2016. "The age of post-truth politics." *New York Times*, Aug. 24.
- Driver, Julia 2003. "The Conflation of Moral and Epistemic Virtue." In *Moral and Epistemic Virtues*, ed. by Michael Brady and Duncan Pritchard. Oxford: Blackwell.
- Economist*. 2017. "The Politics of Outrage," and "Politics and Social Media," Nov. 4, 11, 21-24.
- Fandos, Nicholas. 2017. "White House Pushes 'Alternative Facts.' Here are the Real Ones." *New York Times*. Jan. 22.
- Farrell, Joseph and Matthew Rabin. 2006. "Cheap Talk." *Journal of Economic Perspectives*, 10:3, 103-118.
- Fisher, Marc. 2017. "The Political Lexicon of a Billionaire Populist." *Washington Post*. March 9.
- Gershman, Samuel et al (2015). "Computational rationality: A Converging Paradigm for Intelligence in Brains, Minds, and Machines." *Science*, 349, 273-278.
- Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.
- Grice, H. P. 1975. "Logic and Conversation." In *Syntax and Semantics 3: Speech Acts*, ed. by Cole et al. Elsevier.
- Grice, H. P. 1969. "Utterer's Meaning and Intention." *Philosophical Review*, 78:2, 147-77.

- Grice, H. P Paul. 1957. "Meaning." *The Philosophical Reivew*. 66:3, 377-388.
- Guala, Francesco. 2006. "Has Game Theory Been Refuted?" *Journal of Philosophy*, 103:5, 239-263.
- Habermas, Jurgen. 1984. *The Theory of Communicative Action*, vol. I. Trans. by Thomas McCarthy. Boston: Beacon Press.
- Hausman, Daniel M. 2011. *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.
- Heath, Joseph. 2001. *Communicative Action and Rational Choice*. Cambridge: MIT Press.
- Hubin, Donald C. 2001. "The Groundless Normativity of Instrumental Rationality." *The Journal of Philosophy* 98:9, 445-465.
- Ibrahimi, Khalil, Eitan Altman, and Majed Haddad. 2013 "Signaling game-based approach to power control management in wireless networks." *Proceedings of the 8th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks*. ACM.
- Keyes, Ralph. 2004. *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. St. Martin's Press.
- Lewis, David. 1969. *Convention*. Oxford: Blackwell.
- Maheshwari, Sapna. 2017. "20th Century Fox Gives Real Apology for a Fake News Campaign," *New York Times*, business, Feb. 16.
- McCloskey, Deirdre. 2015. *Bourgeois Equality*. Chicago: U. Chicago Press.
- Moore, G. E. Moore. 1953. *Some Main Problems of Philosophy*. London: MacMillan.
- Nemeth, Charlan. 1972. "A Critical Analysis of the Research Using the Prisoner's Dilemma Paradigm for the Study of Bargaining." *Advances in Experimental Social Psychology*, vol. 6, 203-234.
- Poovey, Mary. 1998. *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. University of Chicago Press.
- Searle, John R. 2001. *Rationality in Action*. Cambridge: MIT Press.
- Shapin, Steven, and Simon Shaffer. (1985). "Leviathan and the Air Pump: Boyle, Hobbes, and the Experimental Life."

Shoham, Yoav, and Kevin Leyton-Brown. 2008. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge U. Press.

Sobel, Joel. (2012) "Signaling games." *Computational Complexity*. Springer New York, 2012. 2830-2844.

Stalnaker, Robert. 2006. "Saying and Meaning, Cheap Talk and Credibility." In Benz, Anton and Gerhard Jäger, Robert van Rooij. 2006. *Game Theory and Pragmatics*. ISBN: 978-1-349-52317-7 (Print) 978-0-230-28589-7 (Online)

Turner, Stephen P. 2010. *Explaining the Normative*. Cambridge UK, Polity.

Williams, Bernard. 1997. "Moral Responsibility and Political Freedom." *Cambridge Law Journal* 56.1: 96-102.

Wright, Kevin B. and Lynne M. Webb, eds. 20xx. *Computer-mediated Communication in Personal Relationships*.

Zollman, Kevin J.S., Carl T. Bergstrom and Simon M. Huttegger. 2012. "Between Cheap and Costly Signals: the Evolution of Partially Honest Communication." *Proceedings of the Royal Society B*. Nov. 7.

Table 1

		Receiver	
		“One”	“Two”
Sender	“One”	1, 1	-1, -1
	“Two”	-1, -1	1, 1

Table 1

Table 2

		Ralph	
		Cooperate	Defect
Sarah	Cooperate	50, 50	0, 100
	Defect	100, 0	25, 25

Table 2

Table 3

		Employer	
		High	Low
Candidate	High	2, 1	0, 0
	Low	2, 0	1, 3

Table 3

Table 4

		Receiver	
		L	T
Speaker	R	10, -10	-10, 10
	NR	0, 2	0, -2

The payoff refers to the outcome in the case that it is either raining or not raining. The speaker chooses to send the signal “It is raining” [R] or “It is not raining” [NR] and if the receiver finds the signal credible takes the appropriate action of leaving [L] her umbrella if it is not raining, and taking [T] her umbrella if it is raining. If the receiver leaves her umbrella, the speaker can take it. From this game payoff matrix, the receiver will know that the speaker receives no positive reward for inducing in her the belief that it is raining, loses value if she believes it is raining when it is, and gains the most value if she is induced to incorrectly believe it is not raining when it is.

Table 4: Umbrella game