1 Intention Recognition as the Mechanism of Human Communication

Daniel W. Harris

1. The Intentionalist Program

My topic is a research program variously called "intentionalism", "Grice-anism", and "intention-based semantics". Following Grice (1957, 1968, 1969), contributors to this program have sought to show how facts about meaning and communication boil down to facts about our psychology, with a special emphasis on our capacity for revealing and recognizing intentions.

The central claim of intentionalism is that what someone means when speaking or otherwise attempting to communicate is a matter of what they communicatively intend. An agent communicatively intends something when they make an utterance with the intention of having a certain effect on their addressee's mind, partly by way of revealing to them the intention to do so. Communication happens when the addressee recognizes what kind of effect the agent intends to have on them. An "utterance", in the sense at issue here, can be any overt behavior that serves as evidence of the agent's intentions, and needn't be verbal or linguistic. In the right circumstances, one can communicate via an idiosyncratic gesture, by making a U-Turn on one's scooter, or by any other observable action. What language adds to this picture is a powerful way of offering rich and systematic evidence of what one means.

In the work of Brian Loar and Stephen Schiffer, intentionalism became part of a research program whose explicit aim is to show how meaning fits into the natural order.[1] One thing that makes this a pressing issue is that the kind of communicating that we do—particularly, but not only, with language—is hard to come by elsewhere in nature. A few other species of animals are capable of formulaic noises, gestures, or dances that prompt their conspecifics to avoid threats and seize opportunities in their immediate environments. But their simple signaling systems are so much less powerful and flexible than the distinctively human variety that the latter almost certainly demands a *sui generis* explanation.

The quest to naturalize meaning would have seemed particularly pressing in the mid-20th century, when philosophy still suffered from the hangovers of behaviorism and positivism and when semantic vocabulary

was met with distrust. Loar, Schiffer, and others conceived of intentionalism as part of a larger reductionist project: first use a broadly Gricean strategy to reduce facts about what words and sentences mean to facts about what speakers mean by them, then reduce facts about what speakers mean to facts about their psychological states, and then find some way to further reduce the needed parts of psychology to more basic domains—ultimately to physics.

My focus here will be on the part of this project that belongs to the philosophy of language. What are the prospects for a reductive explanation of the facts about human communication, including linguistic communication, in terms of intention recognition and related psychological capacities? In particular, how should this research program be updated in light of the theoretical and empirical resources now available?

2. Intentionalism and Explication

Understood as a reductive program, intentionalism seeks to explain higher-level phenomena to do with meaning in terms of lower-level phenomena to do with psychology and ultimately physics. How is this talk of "levels" and "explanation" to be understood?

Beginning with Grice, the central hypotheses of intentionalists have often taken the form of explications such as (1).

1. "*U* meant something by uttering *x*" is true iff, for some audience *A*, *U* uttered *x* intending

    (i) *A* to produce a particular response *r*
    (ii) *A* to think (recognize) that *U* intends (i)
    (iii) *A* to fulfill (i) on the basis of his fulfilment of (ii).

    (Grice 1969, 151)

Although these explications are often stated with symmetric connectives like "iff" or "$=_{df}$", it is clear that something asymmetric is always intended. The schematic statements on the left are there to have light shed on them, and those to the right are there to do the shedding. The right is the *explicans* to the left's *explicandum*. The left is a high-level description of the same phenomenon that the right describes at a lower level.

What is the point of explication? One traditional answer is that explications offer analyses of concepts. On this view, (1) is an attempt to articulate what is already there in one of our ordinary concepts of meaning. It is an attempt to unpack the concept's underlying structure in terms of more fundamental concepts.

Some work in the intentionalist program—particularly most early work—makes sense only if we think of explication as something like conceptual analysis. Grice begins "Meaning" (1957) by distinguishing

natural and non-natural meaning before subdividing the latter into time-less meaning and utterer's meaning. His method is to attend to subtle cues about our ordinary usage of "means" and its cognates. "I cannot say 'those spots meant measles, but he hadn't got measles'", Grice tells us, but one *can* say, "'Those three rings on the bell . . . mean that the bus is full' . . . and go on to say, 'but it isn't in fact full—the conductor has made a mistake'" (Grice 1957, 377–378). The "can" here is that of felicitous usage. As competent users of the English verb(s), "to mean", we are intended to knowingly agree. Grice goes on to offer a tentative explication for each kind of non-natural meaning. In later work, Grice and others would painstakingly hone these explications by means of the method of cases. Particularly when it came to utterer's meaning (a.k.a. "speaker meaning"), an enormous number of explications have been proposed, and each has been met with intuitive counterexamples.[2] These counterexamples take a standard form. A situation is presented in which someone meets the proposed criteria for utterer's meaning but intuitively fails to instantiate the concept under investigation (or vice versa). As Grice (1969, 155) puts one such *coup de grace*, "I do not think that one would want to say that *U* had meant something by throwing the banknote out of the window". Again, we are meant to nod along: the key premises of these arguments are judgments about how we, ordinary speakers, would use the word "mean". Given this methodology, in which an explication's success hangs on its accordance with ordinary usage, the aim of explication can only be to analyze our concepts or to spell out the meanings of our words.

This project is methodologically ill conceived. One reason for pessimism is inductive. Since Plato, philosophers have been trying very hard to give necessary-and-sufficient conditions for some concepts in terms of others in a way that stands up to intuitive scrutiny. So far, we have succeeded approximately zero times. This is not conclusive proof that the job can't be done, any more than the failures of alchemists prove that lead can't be made into gold. But it should not inspire confidence.[3] Another reason for pessimism arises from the broader goals of intentionalism. We seek to boil meaning down to psychology and ultimately to physics. But psychologists and physicists discover new things all the time, and they sometimes even engineer new concepts in which to frame these new things. Most of us therefore lack some of the concepts (and most of the truths) in terms of which statements about meaning would have to make their way down the conceptual hierarchy. Our intuitions about counterexamples cannot bear the epistemic load of this procedure.

By the early 1980s, we find Loar and Schiffer explicitly distancing themselves from conceptual analysis. Loar (1981, 41–43) tells us that "philosophical explication does not have to be seen as Moorean analysis", argues that it needn't "depend on any non-natural analyticity", and gestures at explication as a naturalistically respectable practice in the

philosophy of science. According to Loar, the aim of an explication is to conservatively replace ordinary concepts with concepts capable of doing serious theoretical work.

Schiffer (1982, 120) is even more explicit about his methodological aims:

> Certain intention theoretical writings have, unwittingly, tended to foster the misleading impression that the program was an exercise in conceptual analysis, the aim and the end of which was the definition of various ordinary language semantical idioms in terms of certain complexes of propositional attitudes. . . . In fact, the program need have no truck with conceptual analysis. . . . The intention theorist seeks to reduce the having of content of marks and sounds to the having of content of psychological states.

How, then, to proceed? Schiffer retains his own earlier explications, but adorns the central terms of his explicanda with asterisks—"meaning*" instead of "meaning", and so on—in order to mark them out as technical terms rather than items from our ordinary vocabulary. We should understand the explications as "stipulative definitions" which "may not be realized" (1982, 123). Once thereby linked to lower-level claims, the top-level claims serve as empirical hypotheses about what people are doing when they speak. We all know that speakers mean things by their words. But do they also mean* things by their words, with all that that (stipulatively) entails? Schiffer's claim—an empirical hypothesis—is that they do.

This methodology is a clear improvement over the idea that intentionalism is a project in conceptual analysis. It allows us to be in the business of discovering the hidden underlying nature of meaning and communication, rather than merely reformulating what is supposedly already implicit in our conceptual schemes. It gives sense to the idea that our reductive project ultimately aims to recruit parts of science that most of us don't know much about. And it relieves us of the misguided task of testing explications against our intuitive judgments about cases.

Or, at least, it should. The Schiffer of the 1980s still appeals to our judgments about cases—for example, when arguing that Grice's explication of utterer's meaning works better as an explication of "telling" (Schiffer 1982, 121). There is also something fishy about the fact that his explications, presented in this new methodological context, are the very same explications that he'd painstakingly arrived at a decade earlier via the method of cases. Given our rejection of conceptual analysis and, with it, the central methodological role played by intuitive counterexamples, aren't these explications ill-gotten gains? Shouldn't we have to begin from scratch with careful, empirically-grounded theorizing about what kind of psychological states and processes explain our communicative capacities?

3. Intentionalism and Mechanistic Explanation

Latter-day intentionalists have mostly eschewed the explication-centric methodology of Grice, Loar, and Schiffer. In part, this is because the intentionalist program has migrated from philosophy into cognitive science. Some have focused on developing models of the cognitive processes by means of which humans recognize communicative intentions.[4] Some have investigated the processes by which we design our utterances to affect particular addressees' mental states.[5] Others have focused on the interrelated development of intention-recognition, communication, and language in children.[6] Still others have investigated the role of intention-recognition in the evolution of language and communication, in part through comparative study of human and non-human communication.[7] This work belongs to the intentionalist program. But its methodology differs from earlier work. One finds few explications in the style of (1), and even fewer intuition-driven counterexamples. In their place is a wide range of empirical considerations, all bearing on the extent to which our communicative capacities can be explained in terms of our underlying capacity for intention recognition. Although we are still dealing with a kind of reductive explanation, it is not of the kind pursued by Loar and Schiffer.

I think that we can make sense of this shift in the methodology of intentionalists by considering some recent work by philosophers of science on the nature of reductive explanation in the special sciences more generally.

Consider the model of inter-theoretic reduction that was, until recently, dominant in the philosophy of science. Suppose we have a high-level theory that is framed in a proprietary stock of predicates and we want to reduce it to a lower-level theory with its own predicates. To do this, we formulate *bridge principles* that, together with the laws of the low-level theory, entail the laws of the high-level theory (Nagel 1961, ch.11). These bridge principles take the form of biconditionals that connect the two theories' vocabularies. By thus reducing the high-level theory to the low-level theory, we explain the truth of the former in terms of the more fundamental and better-established truth of the latter.

A lesson of recent work in the philosophy of science is that although this sort of account may make sense of reductive explanations in physics—for example, the reduction of thermodynamics to statistical mechanics (Nagel 1961, 338–44)—it mischaracterizes the kinds of successful reductive explanations that one finds in biology, neuroscience, psychology, and other high-level special sciences. One big problem with the traditional model is that formulating predictive laws isn't among the central activities of these sciences, and so it is not possible to show how such laws stand in entailment relations to laws at higher or lower levels.

A wave of recent work has argued that the dominant form of reductive explanation in biology and related special sciences involves showing

how processes at one level are explained by the activities of underlying *mechanisms* at lower levels.[8] Godfrey-Smith (2008, 54) summarizes recent thinking about this style of explanation as follows:

> When we look at successful reductionist research programs in areas like biology, we do see an accumulation of information about how various biologically important processes occur. We now have a good understanding of processes like photosynthesis, respiration, protein synthesis, the transmission of signals in the brain, the action of muscles, the immune response, and so on. This sort of work can reasonably be, and often is, described as reductionist. We are taking a high-level process or capacity, and explaining how it works in terms of lower-level mechanisms and entities.

In mechanistic explanations, relations between levels of explanation are part-whole relations rather than entailment relations. In molecular biology, for example, reductive explanation usually amounts to identifying the distinct physical parts of a larger system and showing how their chemical and mechanical activities add up to the activities and the system as a whole. To explain the transmission of chemical energy through synapses, for example, we identify the parts of neurons that contribute to this process and show how their activities contribute to the larger process in which they play a role (Machamer, Darden, and Craver 2000, 8–13).

More recently, the idea of mechanistic explanation has been extended to characterize reductive explanation in cognitive science.[9] In this context, the mechanisms in question may be subsystems that are individuated in terms of their information-processing roles within a larger system, rather than as physical parts of an entity (Bechtel 2008, ch.1). Still, it makes sense to explain the activities and capacities of a system in terms of the capacities, activities, and interactions of its subsystems. This is just the project of investigating human cognitive architecture, which has been a central task in cognitive science for decades.

My point in summarizing these methodological ideas is that I think that they should be self-consciously absorbed by the intentionalist program. Like the domains that have interested the new mechanists, our explananda are not well understood in terms of scientific laws or even well-organized theories. There are no laws of speaker meaning or communication. It is a bit strange even to say that we have a theory of these things that we are seeking to reduce. Even semantics, which is now a mature scientific research program, does not produce laws but a body of lexical semantic values and composition rules, mental representations of which are thought to play a causal role in language processing.

What intentionalism seeks to explain is the human capacity to communicate. Our method for doing this should be to study the systems and subsystems of the mind that are at work when we mean things by our

utterances and work out the meanings of others' utterances. We should study how these systems work, how they collaborate, why they sometimes fail, how they evolved in our ancestors, and how they come online in children. In other words: we should study the nature and origins of the cognitive architecture underlying human communication.

This methodology is already embodied, if not explicitly avowed, in some of the cognitive-scientific contributions to intentionalism that I mentioned at the start of this section. There are even glimmers of it in some work of Grice, Loar, and Schiffer. I will point to some of these glimmers in §7, but I will also argue that they have tended to go undeveloped, in part because of an excessive and sometimes competing focus on explication. Although we needn't eschew explications entirely, I think that we should give them a secondary role in our theorizing.

Here is what I propose to do in the rest of this paper. First, in §4, I will outline an intentionalist answer to the question of what goes into our capacity to communicate. In §5, I will sketch a partial theory of the cognitive architecture underlying communication. In later sections, I will put these ideas to work, both by showing how they settle issues within the intentionalist program (§§6–7) and by showing how they dissolve certain objections to it (§8).

4. What Makes Up Our Capacity to Communicate?

What about us makes human communication so much more flexible and powerful than communication in other species? I will focus on three psychological capacities that together go a long way toward explaining what is unique about us. These are *mindreading*, *planning*, and *language*.

To mindread is to treat a part of the world as an agent and predict or infer its mental states. Humans have a uniquely powerful capacity and compulsion to mindread. We continually track the thoughts and motives of the people around us, and our go-to strategy for understanding their behavior is by reasoning about their states of mind. Mindreading has been one of the most intensively researched topics in cognitive science since the 1980s, spanning developmental and cognitive psychology, cognitive ethology, and the philosophy of psychology.

To identify a communicative intention is to engage in an advanced form of mindreading. It is to infer the intention as the best explanation of an utterer's behavior, in light of the utterance they've produced and whatever other background information may be available and relevant. Mindreading also plays a crucial role in the performance of communicative acts, since designing an effective utterance means predicting how it will impact the mind of one's addressee.

By planning, I mean the process of forming intentions and reasoning from those intentions together with one's beliefs to further intentions, all for the purposes of coordinating one's actions.[10] Planning has been a

central topic in the philosophy of action since the seminal work of Bratman (1987), who emphasized that serving as the premises and conclusions of practical reasoning is central to the functional role of intentions.

One part of planning is conceiving of options and choosing between them by weighing preferences and emotional reactions. This is a process that decision theorists, psychologists, and neuroscientists have long studied. One of Bratman's central insights was to recognize that a theory of choice would capture only one narrow step in the planning process. This is because the way that an agent makes any given choice is constrained by the intentions they have formed as a result of prior practical reasoning. To choose is normally to select a subplan of one's prior plan, a way of implementing intentions that one already has. Intentions are relatively stable commitments that play a dual role of controlling action and constraining the options that must be considered in further practical reasoning. Planning is a way of overcoming our cognitive limitations by letting prior decisions lessen the burden of making further choices.

Our capacity for planning is part of our capacity for communication because communicative intentions function in the usual way as both the premises and conclusions of practical reasoning. Each communicative intention is itself a subplan of our other prior intentions, the result of practical reasoning about the proper means to our broader ends. If what I communicatively intend when I say "buy me a drink" is for you to form an intention to buy me a drink, I do so not for no reason, but presumably as a means to the end of getting a drink. Each communicative intention itself has the structure of a complex plan: my intention for you to recognize what kind of effect I intend to have on you is intended to be a partial means to the end of having that very effect. This is what sets communicative intentions apart from less cooperative ways of getting people to change what they believe and intend, and it is what allows us to use utterances that give evidence of our intentions as a means of realizing those very intentions.

Once a communicative intention is formed, it must serve as a premise in further reasoning about what kind of utterance to make. If you and I have a habit of buying each other drinks, I may be able to get my point across by simply pointing to the bar and miming the action of tilting a glass toward my lips. If I can't rely on this background, I might have to be more explicit: "How about if you get me an IPA for this round and I'll get the next?". Choosing between these and the many other possible ways of getting you to recognize my intended effect depends on fast and intelligent practical reasoning that draws on both my prior plans and what I know about you, including what I am able to infer about your actual and potential mental states.

What I have said so far applies to nonlinguistic as well as linguistic communication, which are both, at their core, driven by planning and mindreading. What distinguishes the two is that linguistic utterances are

the product of a capacity that allows us to encode and decode powerful and highly structured evidence of our intentions. (I will say more about this in §5.)

The distinctive claim of intentionalism, as I conceive of it, is that intention recognition is the process at the core of distinctively human communication, and that language plays a subservient role when it is involved at all. The most explicit articulations of this claim in the literature have been due to Sperber and Wilson (1995, ch.1) and Scott-Phillips (2014, ch.1), who make the point by distinguishing two models of communication. In the "code model", a sender encodes some informational content in a signal and sends it through a transmission channel to a receiver, who decodes the signal to access the content. In the "ostensive-inferential" model, communication is a process in which the receiver must combine the evidence gleaned from the sender's utterance with other information at their disposal in order to infer the sender's intention. The two models may be combined when an encoding—decoding process subserves intention recognition. Following Sperber and Wilson, I claim that this is what is going on when we communicate with language: a speaker encodes evidence of what they communicatively intend in a linguistic utterance, but merely decoding this utterance is never sufficient on its own for successful communication; further inference is always required to identify the speaker's intention.

Intentionalism is usually associated with the idea that speaker meaning—and so, intention recognition—takes explanatory priority over linguistic meaning. There is a clear sense in which I agree with this claim: I think that linguistic processing always subserves intention recognition in human communication. This is a claim about the priority of one process over another in actual episodes of communication.

There are other priority claims that might interest us as well. Some have argued that language evolved to subserve communication by intention recognition, or that the biological function of language is to play this role.[11] Similarly, some have argued that mindreading plays a central role in certain aspects of language acquisition—a view on which at least some of the mechanisms underlying intention recognition are developmentally prior to the capacity for language.[12]

These are interesting claims, and if they turn out to be true, then they lend strong support to intentionalism. But I want to stress that intentionalism, as I have articulated it, does not entail these evolutionary or developmental claims, and can be true even if they are false. It could be that our language mechanism(s) evolved to play some non-communicative role but were later repurposed for communication. Likewise, it might be that the biological function of the mechanisms underlying our language capacity is not, or is not *only* to subserve communication, but that we routinely use it to communicate anyway. (By analogy, typing on a keyboard would not be the biological function of human fingers even in a

near-future dystopia in which that's all we use them for.) Similarly, it could be that language acquisition is accomplished almost entirely by domain-specific mechanisms and does not recruit the mechanisms underlying intention-recognition. But this independence in acquisition is compatible with my claim that language subserves intention recognition in communication. So, although intentionalism has often been presented as having evolutionary and developmental components, and although I find those ideas to be interesting and worth pursuing, I wish to separate my central claim from them here.

5. Communication and Cognitive Architecture

How are our capacities for mindreading, planning, and language realized in our cognitive architecture? In previous work, I have defended a partial answer to this question that draws on a model of the mind originally developed by Fodor (1983). On Fodor's view, the mind is divided into one or more central-cognitive systems and an array of peripheral input-output systems that mediate between central cognition and our sensory systems (in perception) and motor systems (in action). These input-output systems, or modules, are set apart by a cluster of features: they are fast, automatic, domain-specific, functionally dissociable from other systems, and in some cases associated with specific brain areas or evolutionary origins. The trait that Fodor has most emphasized is that modules are informationally isolated from each other and from central cognition. Although they send outputs to and/or take inputs from central cognition, their internal processes are insensitive to beliefs and memories to which central-cognition has access. Instead, they draw on proprietary databases of information to which central-cognition in turn lacks access. This allows them to perform their tasks quickly and in a way that generally isn't under voluntary control. Because their informational resources are circumscribed, modular systems are amenable to computational modeling in a way that abstracts from other cognitive resources. This sort of computational modeling tends to be less effective when it comes to what Fodor thought of as central-cognitive systems, which are less restricted in the information on which they can draw.

One of Fodor's main examples of a modular input-output system is the language system. On Fodor's view, linguistic theory is the study of the databases on which linguistic processing draws, and psycholinguistics is the study of how these databases develop and are deployed in the perception and production of speech. In recent work, I have argued that the language system includes a semantic subsystem, and that our best current semantic theories should be understood as models of this subsystem's proprietary database of information, on which it draws in order to bridge the gap between syntactic representations to semantic representations in both the perception and production of utterances (Harris 2017, forthcoming).

In language perception, the inputs to semantic processing are LFs, which are the outputs of syntactic processing. The outputs of semantic processing, I claim, are incomplete representations of the communicatively intended effect associated with the sentence being perceived. When I hear someone utter "he lives in Montana", my semantic system outputs a representation of a property shared by beliefs that predicate living in Montana of some male individual. This representation is passed to the central-cognitive system responsible for mindreading, where it serves as partial and defeasible evidence of the kind of mental state that the speaker intended to put me in. In this case, my mindreading system would have to infer which male is being referred to, whether the speaker is being wholly literal, and, if not, what else they might mean.

In language production, the system responsible for planning sends instructions to the language system, which transforms these instructions into phonological representations of an utterance, which are then turned into motor instructions for the relevant articulatory system. I will refer to the instruction that the planning system sends to the language system as an *utterance plan*. I assume that utterance plans are themselves intentions. When one produces an utterance in order to communicate, one's utterance plan is itself a subplan of one's communicative intention. Language production has been much less studied than language perception, and there are difficult questions about how each step in this process works. I will bracket most of these questions here. In particular, it is unclear just what information is included in utterance plans.[13] Suffice it to say that utterance plans must include some information about the mental attitude and content of a mental state to be produced, as well as some other crucial information about an addressee and a context. For example, if the content to be communicated is about the addressee, then the language system will need this information so that it can select a second-person pronoun rather than a third-person pronoun in the appropriate position.

In brief, my argument for the modularity of semantics rests on explanatory inferences from four empirical premises.[14] First, contemporary semantics has had to painstakingly reverse engineer all that we know about expressions' semantic values and the principles by which they compose. The best explanation is that this information is inaccessible to central cognition, and so unsusceptible to introspection. Second, semanticists have had a lot of success with this project, and have been able to generate precise empirical predictions about sentence meanings from concise semantic theories. By contrast, we have no predictive theory of the processes underlying mindreading. The best explanation for this is that semantic processing, unlike central cognition, depends on an encapsulated body of information that can be concisely modeled. Third, our beliefs, desires, and intentions seem to have little impact on how we compose the meanings of the sentences we perceive. For example, strongly held false beliefs about semantics do nothing to interfere with one's semantic processing ability. The best explanation is that the information drawn on

by semantic processing is largely sequestered from our beliefs and other central-cognitive states. Fourth, contemporary semantics tells us that the principles that guide semantic processing are framed in terms of concepts like *semantic type*, *numerical index*, and *assignment function*, but most competent speakers lack the ability to form beliefs and other central-cognitive representations in terms of these concepts. The best explanation of this is that the semantic module has its own proprietary conceptual repertoire, which only partially overlaps the conceptual repertoire of central cognition. These arguments give us a compelling reason to think that the body of information that underlies a speaker's semantic competence is cut off from their beliefs, intentions, and other central-cognitive representations, and that it is this body of information of which semantics is the study.

My position that semantic processing is subserved by a subsystem of a modular language system leaves open a range of positions about the cognitive architecture underlying mindreading and planning.

One possibility is that mindreading and planning are not the products of distinct subsystems at all, but are both grounded in a general-purpose central-cognitive system. Some have read this view into Fodor (1983), who points out that there seem to be no principled limits to the kinds of beliefs, intentions, or other central-cognitive representations that may, on occasion, factor into mindreading and planning. Some psychologists have likewise argued, mostly on developmental grounds, that mindreading is an application of our general-purpose cognitive capacity for theorizing about the world.[15]

On the other hand, some have found reasons other than informational isolation to posit specialized mindreading and planning systems. Unlike various other kinds of theoretical knowledge, mindreading is universal to cognitively unimpaired humans, develops on a relatively fixed and predictable schedule, and is selectively impaired in people with autism (Baron-Cohen, Leslie, and Frith 1985; Baron Cohen 1997) and in people with damage to certain brain areas (Rowe et al. 2001; Frith and Frith 1999). These considerations have led some to posit one or more innate, domain-specific mechanisms and/or bodies of knowledge underlying the mindreading capacity.[16] Some have argued on similar grounds that planning is the product of one or more dedicated cognitive mechanisms[17] And some have argued that the systems underlying planning and mindreading are deeply intertwined, on the grounds that mindreading recruits the planning system in order to simulate others' planning processes.[18]

Although I find some of the arguments for discrete systems underlying mindreading and planning persuasive, I won't take a stand on these issues here. I will sometimes speak of "the planning system" or "the mindreading system", but these phrases should be understood to be compatible with the possibility that mindreading and planning are performed by the same system, or by some perhaps-overlapping constellation of systems.

6. Semantics Doesn't Reduce to Propositional Attitudes

My position on the informational isolation of semantics is incompatible with the view that the semantic facts about natural language can be wholly explained by speakers' propositional attitudes—a view that has often been endorsed by intentionalists. This view is suggested, for example, by Grice's (1957, 385) original, albeit thoroughly hedged explication of utterance-type meaning:

> "x means$_{NN}$ (timeless) that so-and-so" might as a first shot be equated with some statement or disjunction of statements about what "people" (vague) intend (with qualifications about "recognition") to effect by x.

Loar and Schiffer attempted to concretize Grice's proposal by combining it with Lewis's (1975, 1969) theory of linguistic convention, which itself explicates conventions in terms of participants' common knowledge about their regularities of belief and behavior. As a result, we find Loar and Schiffer saying things like this:[19]

> If we knew all the communicative intentions and other propositional attitudes of members of P, as well as their correlations with utterances of sentences of L within P, and if we had time enough and computational power for ideal reflection, we could then directly infer a priori that L is the language of P.
> (Loar (2001); reprinted in Loar [2017, 127])

> once we have decided on a scheme of attributing propositional attitudes to a population—and that is presupposed, in Lewis's theory and in mine, in assigning a language to a population—then there is a non-arbitrary way of selecting one grammar.
> (Loar 1976, 160)

Loar follows Lewis in taking a language to be a pairing of sentences with meanings and a grammar to be a finite specification of a language. If population members' propositional attitudes determine the language they use, then they also determine the meanings of all of the sentences of their language. And if their propositional attitudes determine the grammar that we should attribute to them, then all of the semantic properties of their language are thereby determined. But if semantics is the study of a semantic module, as I have claimed, then this is incorrect. The body of information that determines the semantic properties of the language I speak is not what I believe or know, but rather the proprietary database of a mechanism that plays a causal role in the interpretation and production of meaningful speech.

7. Explication Versus Explanation

Intentionalists have sometimes shown recognition of the phenomena that have led me to posit the modularity of semantics. Near the end of *Mind and Meaning*, Loar (1981, 259) argues that "the Chomskyan idea of the internalization of the generative procedures of a grammar has got to be invoked to make sense of the entrenchment of a grammar, and therefore to make sense of literal meaning" and goes on to say that "the exact force of this can't be spelled out antecedently to a detailed psycholinguistic theory". In saying these things, Loar stands at the edge of an empirically informed inquiry into the actual psychological underpinnings of our linguistic competence.

But he turns back. His top priority is to give necessary-and-sufficient conditions for the meaningfulness of expressions. And although what actually explains our semantic competence with a language may be certain contingent features of our psychological makeup, Loar judges that these aren't essential features of users of a language like ours. Two users of the same language could have different psychological makeups, he reasons. He therefore goes on to say that his view ultimately does not depend on "whether a psycholinguistic theory of such an adventurous kind is true" (Loar 1981, 260).

Elsewhere, Loar (1976, 158) spells out this line of thought more explicitly:

> There may be a Chomsky sense of knowledge—having an internal representation—in which a speaker knows the rules of his language, but that is a psychological hypothesis and, however reasonable it is, we do not want to build it into an *explication* of what it is for *L* to be the language of *P* Better that it should be offered, at a later stage, as an *explanation* of how it is possible for a complex entity like English to be the language of the population of English speakers.

I think that this passage helps us to see the downside of the traditional, explication-driven methodology. In the quest to give necessary-and-sufficient conditions that apply to all actual and possible users of a language, we are driven to prescind from more and more of the actual psychological details of how humans use language to communicate, since there *could* be creatures who outwardly use language like us but whose psychological inner-workings are different.[20] Explication and explanation become competing goals in a situation of this kind; achieving the former forces us to postpone the latter.

Indeed, Loar and (early) Schiffer don't prescind far enough. Their propositional-attitude-laden explications aren't even true of humans—or so I have argued. Something that abstracts even more from psychological detail is needed. Schiffer (2003, ch.3, 2006) addresses this problem in more recent work, where he argues that semantic competence is a matter of having *some* "information-processing state" that plays the "knowledge-of-meaning

role". To play this role, a state must causally mediate between perception of an utterance and a belief that the speaker thereby performed a speech act with a certain kind of content and a certain kind of force. This is a role that could be played by propositional attitudes or, as I have argued, by an information-processing mechanism that is isolated from cognition. Grice (1968) formulates a similar proposal, on which the meaning of an utterance-type is explicated in terms of the "procedures" that language users have "in their repertoires" to use the expression with certain intentions. Grice does not say what it is to have this kind of procedure in one's repertoire, though nothing that he says rules out either a theory involving propositional attitudes or one involving an informationally isolated language system.

Here Schiffer and Grice prescind almost completely from psychological mechanisms. In so doing, they almost completely forsake explanation in favor of explication. In response to the question of how we comprehend speech, the accounts of Schiffer and Grice can tell us only that we possess *some* information-processing state or procedure for doing so.

In the traditional, Nagelian model of inter-theoretic reduction, explication (via bridge principles) is supposed to *yield* explanation by showing us how to reduce a less-established, high-level theory to a better-established, low-level theory. But because they take our linguistic capacities to be multiply realizable, Loar and Schiffer's quest for explication instead leads them away from explanation.

A common response to this kind of dilemma is to abandon the search for necessary-and-sufficient conditions and to settle for sufficient conditions instead. In place of explication-driven reduction, we could seek supervenience relations, grounding relations, or some other asymmetric dependence relation that holds between the semantic facts and the psychological facts that underlie them—in actual humans if not in some hypothetical behaviorally similar beings. In his last work on intentionalism, Loar (2001) makes this move, articulating a "minimal Gricean thesis" as follows: "Facts about speakers' intentions together with certain other psychological factors asymmetrically determine social meaning; the latter conceptually supervenes, asymmetrically, on the former" (Loar (2001); reprinted in (2017, 127)). Given my approach to intentionalism, I take it to be true that the semantic facts supervene on (and are grounded in) facts about human psychology. The interesting question is: *which* facts about human psychology? The best way to answer this question is to engage in the project of mechanistic explanation that I advocated in §3. The right supervenience (or grounding) claims will be little more than summaries of the results of this project.

8. Non-Communicative Language Use

I want to show how the methodology that I have recommended can help the intentionalist program to make some progress. To do this, I will

consider the perennial objection that intentionalism fails to make sense of non-communicative uses of language.[21]

The objection usually takes the following form. Intentionalists tell us that to mean something by an utterance one must intend to change the mental state of an addressee. But we often use language—meaningfully, it seems—in ways that aren't naturally described in these terms. Wilson and Sperber (1988, 213) imagine someone saying, "please don't rain" while looking up at a cloudy sky, or muttering "start, damn you!" at their car. We often talk to our pets and babies, we practice speeches out loud, and we write in diaries. We speak or write with the aim of better articulating inchoate ideas rather than communicating them to others. We seem to engage in inner speech—a kind of thought that happens in the medium of natural language. These all look like non-communicative uses of language. In particular, these seem to be meaningful uses of language in the absence of communicative intentions.

The traditional intentionalist has a few options for how to respond, and needn't apply any one strategy to all cases.

We should sometimes say that the speaker *does* have the communicative intentions in question. When writing a diary, a speaker may intend to communicate with their future self. Confronted with someone who is talking to a cat or muttering at their car, it might be best to say that they are suffering from a momentary lapse of rationality: they are acting in ways that don't make sense given what they believe, as they would likely admit if pressed.[22]

In other cases, we should deny that the speaker really means anything. When practicing a speech one engages in pretense, which must be distinguished from genuine speaker meaning, just as practicing the footwork for a dance routine should be distinguished from dancing (cf. Grice (1969, 174)). When I am writing in order to clarify my thoughts or engaging in inner speech, I am engaged in a different-enough activity than when I am speaking to another person that it is unhelpful to group them together for theoretical purposes.

We should not be surprised if these defensive strategies don't satisfy all skeptics. They can seem ad hoc. For the skeptic who is still wedded to the method of cases, they may in some cases seem question-begging, since they deny intuitions that the skeptic brings to us as evidence. More importantly, these responses can be accused of missing the real force of some of the purported counterexamples. Traditional intentionalists can explain these cases away, but they offer us few positive resources to explain what people are up to in these cases, or why using language would be a good way to do it.

Some of these uses of language seem to pose threats to the basic assumptions of intentionalism—threats that the above responses don't seem to address. The idea that we use language to clarify our thoughts seems to entail that the content of one's utterance is sometimes more

articulate than the intention one has in expressing it. But how could this be compatible with the claim that what one says is fully determined by what one intends in saying it? And we seem to be put into a vicious circle by the claim that we sometimes think in natural language together with the intentionalist claim that the meanings of natural-language expressions can be explained in terms of the contents of our thoughts.

Loar articulates a related worry:

> The strong thesis is then open to the reasonable objection that, intuitively, natural language meaning could be constituted independently of the use of language in communication. There are such phenomena as working out ideas on a word processor, thinking out loud in sentences, engaging in inner speech: can we not conceive of settings in which they stand on their own, independently of communication? And, more basically, there is the possibility that we normally think, at least in part, in a natural language, that our beliefs and intentions depend on having internalized some natural language, perhaps even in a way that is not typically phenomenologically available. I doubt that it can be argued persuasively on philosophical grounds that this makes no sense. It does not seem to be an obvious conceptual truth that such private meaningful uses of language would presuppose, even indirectly, communicative intentions or linguistic manifestations thereof. The strong Gricean thesis is not something I am inclined to defend.
>
> <div align="right">(Loar (2001); reprinted in (2017, 124–125))</div>

This is not an objection to any particular Gricean explication. Rather, it is an objection to the whole intentionalist premise that the concepts of speaker meaning—and so, communicative intentions—will play an essential role in explaining semantic notions. The intentionalist's concepts are tailor-made to explain communicative uses of language. Loar raises the possibility that communication is just one use of language among many, and so does not deserve a privileged position in an explanation of how language works. In particular, since all of the uses of language that Loar mentions seemingly depend on the semantic properties of words and sentences, Loar's line of thought raises serious doubts for the idea that the semantic facts can ultimately be reduced to facts about the use of language to communicate, to the exclusion of facts about other uses of language.[23] This is one consideration that led Loar to abandon the version of intentionalism that he had previously defended and to fall back a weaker thesis that, he thought, "is perhaps not terribly exciting" (Loar 2017, 131).

These are serious objections, but I think that the methodological reorientation I have been urging puts us in a position to answer them.

Suppose we approach the problem with an eye toward the mechanisms underlying our linguistic capacities. Confronted with a non-communicative use of language, we must ask several questions. First, what is the point of using language in this way? What are we doing, and why? Second, how do we do it? In particular, what are the psychological mechanisms whose activities and interactions allow us to use language in this way? Thirdly, does our answer to the second question undermine the aims and claims of intentionalism? In particular, does it show that there are mechanisms underlying language use other than those that the intentionalist must posit to explain communicative language use? It is a problem for intentionalism if we can't explain how humans perform a certain linguistic activity in terms of the same mechanisms that intentionalists posit to explain linguistic communication. Otherwise, it's hard to see what this objection really amounts to.

9. Some Non-Communicative Uses of Language: A Sketch of an Intentionalist Account

Suppose that the picture of linguistic communication that I sketched in §§4–5 is correct. To communicate is to get an addressee to recognize one's intention to produce an effect in them. When we communicate linguistically, we make use of an informationally isolated language system to encode and decode evidence of our communicative intentions. When someone speaks, their language system takes input from their planning system and encodes a partial and defeasible representation of the mental state they intend to produce, then passes this representation along to a motor system, which turns it into an utterance. When someone perceives a linguistic utterance, their language system decodes from it a semantic representation, which they treat as a partial and defeasible piece of evidence about what kind of effect was intended by the utterance.

This makes it sound as though the language system is a special-purpose device for communicating with others. But note some things that I have *not* claimed: (i) that the only thing for which we use the language system is interpersonal communication; (ii) that the biological function of the language system is communication with others; or (iii) that the language system evolved for communication with others. Although I find the last two of these claims to be plausible, I don't want to commit myself to them here. I see no good reason to accept (i) and some good reasons to reject it. What I need, then, is an account of how a system that plays the communicative role that I have posited it to play can also play various other roles as well. There is nothing mysterious about this: a mechanism can do more than one thing. I normally use my oven to cook food, but I could also turn it on in order to warm up my kitchen. My infant daughter likes to use wooden spoons and measuring cups as

bath toys. Similarly, I propose that the language mechanism, as I have already described it, can be used for purposes other than communication.

First, consider what goes into practicing a speech. In this case, my planning system sends the same kind of utterance plan to my language system that it would send if I were actually giving a speech, but it does this as a way of implementing a different kind of plan. When I am actually giving a speech, the utterance plan that my planning system sends to my language system is a subplan of a communicative intention. This utterance plan is the result of means-end reasoning about how to get my point across to my actual addressee (or group of addressees). When rehearsing a speech, I formulate an utterance plan as a subplan of a different kind of intention—an intention to pretend that I am giving a speech to a certain kind of audience. In this case, the representation of the audience that my planning system sends to my language system is a representation of a hypothetical audience that I've imagined up. It is probably a relatively indefinite representation, but this will also be the case when I give a speech to a roomful of people I don't know.

By sending practice speech through my language system in this way, it may be that I can prime my articulatory mechanisms for a smoother and more skillful delivery later on. I can also listen to my own speech, feeding the outputs of my language system back in as inputs. This may allow me to notice problems that were not apparent at the planning stage, and so adjust my utterance plan to be a better means to the communicative ends that I anticipate having later.

Feeding the outputs of my language system back in as inputs probably also has other uses. I write things down and then read them later as a way of storing information outside my head. In the much shorter term, I may remember a phone number long enough to get out a pen and a piece of paper by repeating it out loud to myself over and over. By feeding information back and forth through my language system in this way, I can use it as a form of short-term memory.

Suppose that we sometimes engage in a version of this activity that doesn't involve actually producing an utterance. When I speak, my language system creates a representation of the phonological properties of an utterance, which is then turned into instructions for my motor systems. Sometimes this last step gets suppressed, and the phonological representation is instead fed back into the language system as an input. This process would amount to a kind of inner speech used for short-term memory.

In fact, there is a large body of evidence that an inner-speech mechanism of this kind plays a role in short-term memory. This mechanism, which is sometimes called the "phonological loop" or "articulatory loop", was posited by Baddeley and Hitch (1974) as one component in their influential model of working memory.[24] According to this model, the phonological loop consists of a memory store and a mechanism for either vocal or subvocal rehearsal. The memory store fades quickly but can be refreshed by

means of rehearsal. There is substantial evidence that subvocal rehearsal is performed by at least some of the same mechanisms that we use to produce utterances. Speaking out loud, even in a way that would not tax the memory store (such as repeating a single syllable), substantially interferes with verbal working-memory performance, for example (Murray 1968). There is also substantial evidence that the storage and rehearsal mechanisms deal, at least in part, in phonological representations rather than purely semantic representations.[25] Phonologically similar strings of letters, numbers, or words are more difficult to store in verbal working memory than phonologically dissimilar strings, for example—an interference effect that is usually explained by assuming that representations in verbal working memory are stored in a phonological format, at least in part (Conrad and Hull 1964; A. D. Baddeley 1966b, 1966a).

This line of thought suggests an explanation of how the language system, conceived as a modular input-output system along the lines I have defended, could serve a variety of non-communicative functions related to pretense, verbal practice, inner speech, and memory. Such an account requires only a couple of speculative hypotheses beyond what I have already defended. First, the utterance plans that the planning system sends as inputs to the language system are sometimes subplans of genuine communicative intentions, but are sometimes subplans of other kinds of intentions, such as the intention to pretend to communicate or the intention to hold something in memory for a short time. Second, inner speech depends on a capacity to intercept the outputs of the language system and feed them back in as inputs before they are turned into motor instructions, resulting in a use of language that Baddeley (2007) calls "subvocal-rehearsal". These hypotheses aren't obvious or uncontroversial, but they are plausible, and they dovetail with a venerable tradition of research on the role of linguistic processing in working memory.

What I have said so far does not on its own account for the puzzling ways in which we are able to clarify our thoughts by speaking out loud or writing things down. Why should sending my thoughts out through the language system and back in again be a way of making them clearer?

Here is a way of answering this question that requires only one further speculative hypothesis. Suppose that the instructions sent by planning systems to language systems are sometimes imprecise—the product of thoughts that are, in some sense, unclear. In particular, utterance plans of this kind aren't precise enough to fully determine what kind of utterance to produce. But suppose that the language system is adept at taking imprecise instructions of this kind and making more or less arbitrary choices about how to turn them into an utterance. It has to pick *some* words and arrange them into *some* syntactic structure, after all, and so it chooses options that are within the range of what is specified, without crashing when the utterance plan it has been given does not narrow the options all the way down.

If we grant this hypothesis, then we have the makings of an explanation of how linguistic rehearsal—subvocal, vocal, orthographic, or otherwise—could be a process by which we make our thoughts more precise. Suppose I have an unclear intentional state, σ. I form an intention to clarify σ through speech, and this intention serves as a premise in means-end reasoning to an utterance plan of the kind that might otherwise be a subplan of a communicative intention to produce σ in someone else. Because σ is unclear, this utterance plan underspecifies the utterance to be produced, and so my language system must make some arbitrary choices in designing an output. This output then becomes an input, which my language system turns into a semantic representation—a partial specification of the kind of effect that someone would intend by using the words that my language system chose. In some cases, the result of this process will be a representation of a more precise mental state than I set out to clarify in the first place. And in *some* cases, I might even prefer this clearer mental state to the one with which I began and decide to adopt it as σ' s replacement. If iterated, this procedure could serve as a selection process by which we hone our thoughts over time.

This is just a sketch of the sort of process by which humans could use a modular language system to clarify their thoughts. This sketch also makes sense of at least one way in which humans could be said to think in language, either inwardly or outwardly. It is a process by which we may sometimes transition from one thought to another, sometimes improved thought. In some cases, the causal pathway of this process is wholly contained within our heads. In others, it follows a trajectory out through articulatory mechanisms and back in through sensory mechanisms. We can say these things without positing a language system that is any more flexible or elaborate than what I have posited to explain linguistic communication. We need only say that we sometimes take advantage of this system's processing capacities for purposes other than communication, and that we do so by sending it the usual sorts of instructions—albeit sometimes less-than-fully-clear ones—but for non-communicative reasons.

Of course, the story I have just told *is* speculative. In particular, a good deal more needs to be said about what it is for one thought to be clearer than another. Although having an unclear thought and clarifying it is something that most of us would say that we regularly do—I am doing it right now!—it is an under-theorized idea.

Still, my aim in sketching this hypothesis has not been to defend it in any detail. I have presented it as an example of how the mechanistic reorientation that I proposed in §3 and implemented in §§4–5 can yield new explanatory resources that give us fresh ways of thinking about old problems. The hypothesis I have offered is promising and deserves to be fleshed out and empirically investigated. It suggests a way of making progress on an issue that has dogged the intentionalist program for over half a century. I, for one, will be pursuing it in future work.

10. Conclusions

I have argued that intentionalism should be pursued as a project in mechanistic explanation rather than as a project in explication. Thus conceived, the research program belongs to interdisciplinary cognitive science rather than to philosophical analysis. Its principal aim is to discern the contours of the cognitive architecture underlying our capacity to communicate, rather than to formulate necessary-and-sufficient conditions for statements about actual and possible communicators.

This methodological reorientation has several advantages. It allows us to turn away from the fruitless pursuit of conceptual analyses and from a quest for necessary-and-sufficient conditions that has led to ever greater abstraction at the expense of explanatory ambition. It puts the intentionalist program in contact with theoretical resources from linguistics, the philosophy of action, and cognitive science that allow us to formulate empirical claims that can be tested and iterated. And it dissolves some objections that have perennially dogged intentionalism, such as the problem of non-communicative language use. Instead of clashing intuitions or a downward spiral of counterexamples, we are left with a scientific research program.

As I suggested in §3, I think that this reorientation has been underway among some intentionalists since the 1980s. At the same time, it remains common to encounter new proposed explications of speaker meaning and new intuitive counterexamples to these proposals. I hope that by explicitly describing and defending my proposed methodological reorientation, I can help us to move toward the future of the intentionalist research program while learning from its past.

Acknowledgements

Notes

1. In particular, see (Loar 1972, 1976, 1981, 2001) and (Schiffer 1972, 1981, 1982, 1993, 2003, 2006).
2. See, especially, (Grice 1957, 1969; Strawson 1964; Schiffer 1972; Neale 1992; Davis 1992, 2003; Sperber and Wilson 2015; Bach and Harnish 1979).
3. Both this point and the snark accompanying it are unoriginal; for more of both, see Fodor (1981, 1998, chs.3–4). The ur-source of skepticism about analysis is Quine (1951).
4. (Sperber and Wilson 1995, 2002; Carston 2002; Wilson and Sperber 2012; Csibra 2010; Keysar 2007).
5. (Clark 1996; Clark and Marshall 1981; Clark and Carlson 1982; Clark and Schaefer 1987; Brennan and Hanna 2009).
6. (Csibra 2010; Hacquard and Lidz 2018; Bloom 2000; Tomasello 2003).
7. (Scott-Phillips 2014; Tomasello 2008; Sperber 2000; Moore 2015).

8. See, e.g. (Machamer, Darden, and Craver 2000; Godfrey-Smith 2008; Bechtel 2008; Bechtel and Abrahamsen 2005; Bechtel and Richardson 1993).

9. E.g. (Bechtel 2008; Godfrey-Smith 2008; Piccinini and Craver 2011). There are, of course, many more-or-less methodologically self-conscious precursors to this approach to cognitive science, and that is part of the new mechanists' point.

10. I use "plan" and "intention" more or less interchangably, but I take "plan" to have the connotation of complexity. As Bratman puts it, "intentions are, so to speak, the building blocks of . . . plans, and plans are intentions writ large" (1987, 8).

11. See, e.g., (Scott-Phillips 2014; Tomasello 2008). For criticisms of these views, see (Chomsky 1976; Bar-On 2013; Hauser, Chomsky, and Fitch 2002).

12. See, e.g., (Bloom 2000; Hacquard and Lidz 2018; Tomasello 2003).

13. I consider the issue in (slightly) more detail in (Harris 2017).

14. For more detailed versions of these arguments and what I take to be their consequences, see (Harris n.d., 2017).

15. (Perner 1991; Wellman 1990; Gopnik and Meltzoff 1997).

16. (Leslie 1987, 1994; Leslie, German, and Polizzi 2005; Scholl and Leslie 1999, 2001; Fodor 1992; Carey 2009; Baron-Cohen, Leslie, and Frith 1985; Baron Cohen 1997; Frith and Frith 1999; Carruthers 2006).

17. E.g. (Carruthers 2004, 2006).

18. (Heal 1986; Gordon 1986, 2009; Goldman 1995, 2006).

19. See also Loar (1972, 1976, 1981, 2001) and Schiffer (1972, 1982, 1993).

20. I confess that I did the same in my dissertation (Harris 2014, chs.4–5).

21. Some influential versions of this objection include those by Chomsky (1976, 55–77) and Carruthers (1996, ch.3). The objection is also one that tends to come up a lot outside print.

22. The human tendency to anthropomorphize inanimate objects is an active topic of psychological research. We tend to anthropomorphize objects when they violate our expectations, including when they malfunction (Epley, Waytz, and Cacioppo 2007). Mandelbaum and Ripley (2012, 364) argue, for example, that we are much more likely to anthropomorphize a car if it doesn't work well. It is a relatively straightforward prediction of intentionalism that phenomena that cause us to anthropomorphize a non-agent will also make us more likely to direct utterances to it.

23. See also Chomsky (1976, 55ff.).

24. For a recent overview of this model of working memory and the evidence that supports it, see (Baddeley 2007).

25. Following Baddeley (2007, 8), my use of "phonological" is not meant to invoke any particular theoretical claims, including those of mainstream phonology. Rather, it is meant to capture the idea that phonological representations include information relevant to articulation in a verbal medium. Others sometimes use the term "phonetic", "acoustic", or "articulatory" instead.

References

Bach, Kent, and Robert M. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Baddeley, Alan. 2007. *Working Memory, Thought, and Action*. Oxford, UK: Oxford University Press.

Baddeley, Alan D. 1966a. "Short-Term Memory for Word Sequences as a Function of Acoustic, Semantic and Formal Similarity." *Quarterly Journal of Experimental Psychology* 18: 362–365.

———. 1966b. "The Influence of Acoustic and Semantic Similarity on Long-Term Memory for Word Sequences." *Quarterly Journal of Experimental Psychology* 18: 302–309.

Baddeley, Alan D., and Graham Hitch. 1974. "Working Memory." *Psychology of Learning and Motivation* 8: 47–89.

Bar-On, Dorit. 2013. "Origins of Meaning: Must We 'Go Gricean'?" *Mind and Language* 28(3): 342–375.

Baron Cohen, Simon. 1997. *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Baron-Cohen, Simon, Alan M. Leslie, and Uta Frith. 1985. "Does the Autistic Child Have a 'Theory of Mind'?" *Cognition* 21: 37–46.

Bechtel, William. 2008. *Mental Mechanisms*. New York; London: Routledge.

Bechtel, William, and Adele Abrahamsen. 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421–441.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

Bloom, Paul. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Brennan, Susan E., and Joy E. Hanna. 2009. "Partner-Specific Adaptation in Dialog." *Topics in Cognitive Science* 1: 274–291.

Carey, Susan. 2009. *The Origin of Concepts*. Oxford: Oxford University Press.

Carruthers, Peter. 1996. *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*. Cambridge, UK: Cambridge University Press.

———. 2004. "Practical Reasoning in a Modular Mind." *Mind & Language* 19(3): 259–278.

———. 2006. *The Architecture of Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Oxford University Press.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Chomsky, Noam. 1976. *Reflections on Language*. Middlesex: Temple Smith.

Clark, Herbert H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, Herbert H., and Thomas B. Carlson. 1982. "Hearers and Speech Acts." *Language* 58(2): 332–373.

Clark, Herbert H., and Catherine R. Marshall. 1981. "Definite Reference and Mutual Knowledge." In *Elements of Discourse Understanding*, edited by A. K. Joshi, B. L. Webber, and I. A. Sag, 10–63. Cambridge, UK: Cambridge University Press.

Clark, Herbert H., and Edward F. Schaefer. 1987. "Concealing One's Meaning from Overhearers." *Journal of Memory and Language* 26: 209–225.

Conrad, R., and A. J. Hull. 1964. "Information, Acoustic Confusion and Memory Span." *British Journal of Psychology* 55: 429–432.

Csibra, Gergely. 2010. "Recognizing Communicative Intentions in Infancy." *Mind and Language* 25(2): 141–168.

Davis, Wayne. 1992. "Speaker Meaning." *Linguistics and Philosophy* 15(3): 223–253.

———. 2003. *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review* 114(4): 864–886.

Fodor, Jerry A. 1981. "The Present Status of the Innateness Controversy." In *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, 257–316. Sussex: Harvester.

———. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

———. 1992. "A Theory of the Child's Theory of Mind." *Cognition* 44: 283–296.

———. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Frith, Chris D., and Uta Frith. 1999. "Interacting Minds: A Biological Basis." *Science* 286: 1692–1695.

Godfrey-Smith, Peter. 2008. "Reduction in Real Life." In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup, 52–74. Oxford: Oxford University Press.

Goldman, Alvin. 1995. "In Defense of the Simulation Theory." In *Folk Psychology*, 191–206. Oxford, UK: Blackwell.

———. 2006. *Simulating Minds*. Oxford, UK: Oxford University Press.

Gopnik, Alison, and Andrew Meltzoff. 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gordon, Robert M. 1986. "Folk Psychology as Simulation." *Mind & Language* 1: 158–171.

———. 2009. "Folk Psychology as Mental Simulation." In *Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, edited by Edward N. Zalta.

Grice, H. P. 1957. "Meaning." *The Philosophical Review* 66(3): 377–388.

———. 1968. "Utterer's Meaning, Sentence-Meaning, and Word-Meaning." *Foundations of Language* 4(3): 225–242.

———. 1969. "Utterer's Meaning and Intention." *The Philosophical Review* 78(2): 147–177.

Hacquard, Valentine, and Jeffrey Lidz. 2018. "Children's Attitude Problems: Bootstrapping Verb Meaning from Syntax and Pragmatics." *Mind & Language* online first: 1–24.

Harris, Daniel W. 2014. *Speech Act Theoretic Semantics*. PhD Dissertation, City University of New York Graduate Center.

———. 2017. "Speaker Reference and Cognitive Architecture." *Croatian Journal of Philosophy* 17(3): 319–349.

———. forthcoming. "Semantics Without Semantic Content." *Mind & Language*.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298: 1569–1579.

Heal, Jane. 1986. "Replication and Functionalism." In *Language, Mind, and Logic*, edited by Jeremy Butterfield, 135–150. Cambridge, UK: Cambridge University Press.

Keysar, Boaz. 2007. "Communication and Miscommunication: The Role of Egocentric Processes." *Intercultural Pragmatics* 4(1): 71–84.

Leslie, Alan M. 1987. "Pretense and Representation: The Origins of 'Theory of Mind'." *Psychological Review* 94: 412–426.

———. 1994. "ToMM, to by and Agency: Core Architecture and Domain Specificity." In *Mapping the Mind*, edited by L. Hirschfeld and S. Gelman, 119–148. Cambridge: Cambridge University Press.

Leslie, Alan M., Tim P. German, and Pamela Polizzi. 2005. "Belief-Desire Reasoning as a Process of Selection." *Cognitive Psychology* 50(1): 45–85.

Lewis, David K. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

———. 1975. "Languages and Language." *Minnesota Studies in the Philosophy of Science* 7: 3–35.

Loar, Brian. 1972. *Sentence Meaning*. PhD Thesis, Oxford.

———. 1976. "Two Theories of Meaning." In *Truth and Meaning: Essays in Semantics*, edited by Gareth Evans and John McDowell, 138–161. Oxford, UK: Oxford University Press.

———. 1981. *Mind and Meaning*. Cambridge, UK: Cambridge University Press.

———. 2001. "The Supervenience of Social Meaning on Speaker's Meaning." In *Paul Grice's Heritage*, edited by Giovanna Cosenza. Turnhout, Belgium: Brepols.

———. 2017. *Consciousness and Meaning: Selected Essays*. Oxford, UK: Oxford University Press.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67(1): 1–25.

Mandelbaum, Eric, and David Ripley. 2012. "Explaining the Abstract/Concrete Paradoxes in Moral Psychology: The Nbar Hypothesis." *Review of Philosophy and Psychology* 3(3): 351–368.

Moore, Richard. 2015. "Meaning and Ostension in Great Ape Gestural Communication." *Animal Cognition* 19(1): 223–231.

Murray, D. J. 1968. "Articulation and Acoustic Confusability in Short-Term Memory." *Journal of Experimental Psychology* 78: 679–684.

Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York, Chicago, San Francisco, Atlanta: Harcourt, Brace & World, Inc.

Neale, Stephen. 1992. "Paul Grice and the Philosophy of Language." *Linguistics and Philosophy* 15: 509–559.

Perner, Josef. 1991. *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Piccinini, Gualtiero, and Carl Craver. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183(3): 283–311.

Quine, W. V. O. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60(1): 20–43.

Rowe, Andrea D., Peter R. Bullock, Charles E. Polkey, and Robin G. Morris. 2001. "'Theory of Mind' Impairments and Their Relationship to Executive Functioning Following Frontal Lobe Excisions." *Brain* 124: 600–616.

Schiffer, Stephen. 1972. *Meaning*. Oxford: Oxford University Press.

———. 1981. "Indexicals and the Theory of Reference." *Synthese* 49(1): 43–100.

———. 1982. "Intention-Based Semantics." *Notre Dame Journal of Formal Logic* 23(2): 119–156.

———. 1993. "Actual-Language Relations." *Philosophical Perspectives* 7: 231–258.

———. 2003. *The Things We Mean*. Oxford: Oxford University Press.

———. 2006. "Two Perspectives on Knowledge of Language." *Philosophical Perspectives* 16: 275–287.

Scholl, Brian J., and Alan M. Leslie. 1999. "Modularity, Development, and 'Theory of Mind'." *Mind & Language* 14: 131–153.

———. 2001. "Minds, Modules, and Meta-Analysis." *Child Development* 72(3): 696–701.

Scott-Phillips, Thom. 2014. *Speaking Our Minds: Why Human Communication Is Different, and How Language Evolved to Make It Special*. Basingstoke: Palgrave Macmillan.

Sperber, Dan. 2000. "Metarepresentations in an Evolutionary Perspective." In *Metarepresentations: A Multidisciplinary Perspective*, edited by Dan Sperber, 117–137. Oxford: Oxford University Press.

Sperber, Dan, and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell.

———. 2002. "Pragmatics, Modularity, and Mindreading." *Mind and Language* 17: 3–23.

———. 2015. "Beyond Speaker Meaning." *Croatian Journal of Philosophy* 15(44): 117–150.

Strawson, P. F. 1964. "Intention and Convention in Speech Acts." *The Philosophical Review* 73(4): 439–460.

Tomasello, Michael. 2003. *Constructing a Language*. Cambridge, MA: Harvard University Press.

———. 2008. *Origins of Human Communication*. Cambridge, MA: MIT Press.

Wellman, Henry M. 1990. *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Wilson, Deirdre, and Dan Sperber. 1988. "Mood and the Analysis of Non-Declarative Sentences." In *Human Agency: Language, Duty, and Value*, edited by Jonathan Dancy, J. M. E. Moravcsik, and C. C. W. Taylor, 210–229. Stanford, CA: Stanford University Press.

———. 2012. *Meaning and Relevance*. Cambridge: Cambridge University Press.