**Intention-based accounts of linguistic communication as idealized scientific models**

Following Grice’s seminal “Meaning” (1957), a diverse array of theories in philosophy, psychology, and linguistics have explained linguistic meaning and communication in part by appeal to interlocutors’ higher-order communicative intentions. Call this feature of these theories *intentionalism*:

**INTENTIONALISM**: A theory is intentionalist if it claims linguistic communication depends on

(a) the speaker's (writer's, signer's) intention to achieve some effect by means of a change in an addressee’s intentional state

and

(b) the addressee recovering meaning by reasoning about the speaker’s communicative intentions

Under this characterization, intentionalists include not only Grice (1957) and later Griceans (Schiffer 1972, Levinson 1983, Horn 1984), but Relevance Theory (Sperber and Wilson 1995), common ground theories in psycholinguistics (Clark 1996, Tomasello 2008) and philosophy (Stalnaker 2002), game-theoretic treatments of meaning (Lewis 1969, Parikh 2000, Clark 2011, Jäger 2012), speech act theories (Austin 1975; Searle 1969), and politeness theory (Brown and Levinson 1987), among other examples. While these accounts differ widely in how they explain linguistic communication, they all share a central commitment to the claim that communication depends on the exercise of *theory of mind*, the ability to reason about the mental states of other creatures.

This leaves an otherwise disparate set of theories vulnerable to the same objection:

---

1 For whom both non-natural meaning and indirect speech acts involve higher-order intentions.
2 Which involves interlocutors considering each other’s desired information and beliefs about the utterance context.
3 In which language use depends on conversational participants’ beliefs about what knowledge they share with their interlocutors.
4 Which models how agents make decisions based on their beliefs about how other agents will make decisions.
5 Which see one component of meaning as the effect of an utterance on the intentional state of the addressee.
6 Which has a Gricean core.
THEORY OF MIND INESSENTIAL (TOMI): there are frequent, non-trivial cases of successful linguistic communication where one or more interlocutors do not exercise theory of mind to produce higher-order communicative intentions. Although controversial, TOMI has the support of multiple lines of empirical research, some of which I will survey in section II. I am less concerned with establishing the truth of TOMI, however, than with determining the intentionalist’s best response. Intentionalist theories include some of the most important accounts of linguistic communication from several disciplines, and we should not give those theories up too quickly.

A number of extant responses to TOMI attempt to reinterpret the data behind the empirical objection in ways more friendly to intentionalism. I will show that the most promising version of this tactic falls short of showing TOMI to be false. We then consider whether the intentionalist can avoid TOMI by loosening the intentionalist criteria. I will argue that this gives up too much, because it abandons the real insights found in appealing to higher-order intentions in communication. A third approach is to limit the aims of intentionalist theories. Most commonly, some philosophers have argued that intentionalism in meant not as a descriptive account of human psychology, but merely as a rational reconstruction, which does not need to answer to the facts in the same way that descriptive theories do. This response also abandons too much. Many intentionalists do have descriptive aims, and it is implausible to see them as presenting rational reconstructions.

The rational reconstruction approach is on the right track, however. Drawing on some comments of Grice’s comments, I will argue that intentionalists are best construed as presenting idealized scientific models. As with a rational reconstruction, the validity of a scientific model is consistent with its falsity. But unlike rational reconstructions, models can aim at accurately representing their target system. Construing intentionalism as a modeling strategy thus allows us to both acknowledge TOMI and preserve intentionalism as a scientifically valid account of linguistic communication.

II. Evidence that theory of mind is inessential to linguistic communication
TOMI has a wide variety of empirical support, and I have space to discuss only some of the more striking examples of research showing that linguistic meaning and communication does not depend on the intentionalist criteria. We will look in detail at three flavors of this objection: linguistic individuals without full theory of mind, communication based on egocentric reasoning, and reflexive, non-intentional communication.

*Children and individuals with non-linguistic impairments*

The first variety of the objection points to a class of people who successfully communicate, but do not seem to be able to have intentions regarding the mental states of their interlocutors. Breheny (2006), for instance, argues that the case of young children is damning for intentionalist theories, specifically targeting Grice, Lewis, Stalnaker, and Relevance Theory. The argument runs as follows:

1. 3-year-olds are competent language users (by developmental linguistic standards).
2. 3-year-olds lack theory of mind (according to developmental psychology).
3. Therefore, theory of mind is not required for linguistic competence.

Since the intentions at the heart of intentionalism involve theory of mind, intentionalism must be false, argues Breheny.

The evidence for (1) is that by age three most children, while not as fluent as adult language-users, frequently produce utterances identical to the utterances an adult would employ in the same circumstances. The evidence for (2) largely comes from the false-belief task (Wimmer and Perner 1983). In the false-belief task subjects are presented with a doll named Sally, who caches a candy in a cupboard. Sally leaves and Anne enters, relocating the candy to a box. Anne leaves, Sally returns, then the subject is asked where Sally will look for the candy first. The correct answer, of course, is the cupboard, but subjects under the age of four consistently fail the task, saying Sally believes it to be in the box. Meta-analysis (Wellman, Cross, and Watson 2001) shows this result to be robust, occurring in most replications and insensitive to whether Sally and Anne are portrayed by dolls or human actors. This is taken to be evidence that young children either do not have or do not employ
theory of mind, and instead use egocentric techniques (i.e. asking themselves “What do I know? rather than “What does Sally know?”) to make predictions about other agents. Thus there is robust evidence that children under four use language without reasoning about their interlocutors’ mental states as intentionalism requires.

A second set of cases used to make the same point are cases of linguistically-competent individuals with an impairment that may render them unable to exercise the requisite theory of mind. Laurence (1996) appeals to these sorts of cases as counterexamples to Lewis’ game-theoretic account of language. The argument is parallel to the argument in the case of young children. Individuals with moderate autism spectrum disorder are often able to communicate linguistically, but consistently fail the false belief task and so purportedly do not have strong theory of mind capacities. In this case as well as the case of young children, we have individuals who seemingly both produce and interpret meaningful utterances without engaging in reasoning about their interlocutors’ mental states. Therefore, their language use cannot depend on higher-order communicative intentions, and intentionalism appears to be false.

Egocentric reasoning in everyday language use

A second variety of empirical evidence against intentionalism claims that the intentionalist criteria are not met even in many cases of everyday, unimpaired, adult linguistic behavior (see, for instance the arguments raised against Tomasello by Peters (2015)). If communication occurs in line with the intentionalist paradigm, then speakers should take into account the mental states of their audience, and vice versa. Some clever experiments, however, seem to show that this is not the case. Keysar et al. (2003) gave subjects the simple task of following instructions given to them by a director, a covert confederate of the researchers. An array of shelves stood between the subject and the director. Some shelves were occluded on one side, so that the subject but not the director could see the contents of that shelf. This meant that reference could differ between the director's

---

7 Though not universally accepted (see section III.)
8 Laurence uses a different example—Williams syndrome—but individuals with Williams syndrome are probably not a good counterexample to intentionalism (Tomasello 1995). Autism spectrum disorder better fills that role in Laurence’s argument.
perspective and the subject's perspective. For example, if the shelves held three candles in three
different sizes, but the smallest was on an occluded shelf, the smallest candle visible to the subject
would differ from the smallest candle visible to the director (see figure 1). Thus, if the subject were
employing theory of mind to interpret the director, she would take the director to mean by "the small
candle" the second smallest candle. If, however, the subject were using egocentric methods to
interpret the utterance, she would interpret “the small candle” to mean the smallest candle,
despite the fact that she should know that the
director is unaware of its existence.

In conditions with indeterminate reference such as the example above, 71% of subjects tried to move the occluded object at
least once, and 46% of subjects did so in more than half of the experimental trials. Variants on
this paradigm involving different sorts of ambiguity, placing the subjects in the role of director, more
naturalistic and free-form tasks, and quantitative eye-tracking and neurophysiological measures have
reliably extended these results (Horton and Keysar 1996, Keysar et al. 2003, Lane et al. 2006,
Apperly et al. 2010, Savitsky et al. 2011).
Keysar and colleagues have argued on this basis that both speakers and addressees frequently fail to consider the intentional attitudes of their interlocutors
during communication. If they are right, this seems to be additional evidence that intentional
reasoning is unnecessary to communication. The claim is not that subjects lack theory of mind
abilities, nor that they never use them in communication, only that sometimes, perhaps most of the

---

9 Keysar et al. were careful to rule out alternative explanations of this failure. Subjects never committed the same sort of
error in a baseline condition, demonstrating that the error must be due to interpretive failure, rather than some other cause
such as motor error. Additionally, in follow-up interviews only one out of thirty-eight subjects reported suspecting that
the director was a confederate, so subjects would not have known that the director was actually aware of the hidden
object.
10 See Brown-Schmidt et al. (2003) for partial dissension.
time, they do not. So long as this occurs with non-trivial frequency, then everyday communication
does not typically depend on higher-order communicative intentions and intentionalism is false.

*Uttering without intending*

The objections to intentionalism we have considered thus far claim that linguistic
communication frequently occurs in the absence of one of the intentionalist criteria. Either
individuals can communicate without theory of mind, or even competent language users often resort
to egocentric rather than intentionalist methods in production and interpretation. The final objection
we will consider goes even further, denying that linguistic communication need involve intentions at
all.

Lind et al. (2014a; 2014b) argue that speech is often reflexive, not intentional. They provide
evidence that our articulatory behavior often occurs in the absence of conscious intention to mean
what we say, and the phenomenology of intention—feeling like we intended to say what we said for
some reason or another—is the result of reflection on our utterances after the fact. In their Stroop
task, subjects were required to produce linguistic responses to the color of a presented stimulus.
Stimuli were printed in colors other than the color denoted by the printed word, and subjects were
instructed to name the color of the text. For example, a subject might see the word ‘blue’ displayed in
red text, in which case the subject should say “red.”

In Lind et al.’s experiment the subject performs the task while wearing headphones. In the
experimental condition, the headphones play a recording of one of the subject’s previous answers
simultaneous to their present response. “Green,” for instance, might play on the headphones as the
subject articulates “gray.” When the onset of the headphone output coincided closely with the onset
of the subject’s utterance, only 32% of subjects suspected some sort of manipulation, and only 4%
were certain that the researchers had played a replacement word over the headphones. Of the
remaining 68% of subjects, 38.5%, when asked, claimed that they had uttered the replacement word.
An additional 16.5% did not need to be asked, because they corrected themselves, saying something
like “no, green” after hearing “gray,” even though they had uttered “green” in the first place. An
additional 29.7%, when asked, did not initially claim to have uttered the replacement word, but revised this answer upon reflection, ultimately admitting to have uttered the word they did not actually utter. Summing these up, 85% of these subjects believed they had said something they had not actually said.

In their analysis, Lind et al. (2014a: 8) suggest that the explanation for these results is that "we actively use feedback to help specify for ourselves the full meaning of what we are saying. In effect, we propose that auditory feedback provides us with a channel for high-level, semantic 'self-comprehension.'" To put it more bluntly, sometimes our perception that we intended to mean $p$ by an utterance is an illusion. We use cues, including hearing our own voice, to determine what we said, and infer that we must have intended to mean $p$ because the effect of what we said was to mean $p$. In reality, however, our linguistic behavior was reflexive, occurring in the absence of such a linguistic intention. Because Lind et al.’s experiment focuses on a common linguistic task—naming—it suggests that reflexive linguistic behavior is common. That their results are more than a fluke is suggested by the similar outcome in Banakou and Slater (2014), who use virtual reality to fool subjects into thinking they both uttered and intended to utter a word actually uttered only by their virtual avatar. Given these experiments it seems that there are cases of meaning which are not the product of conscious intention. Although these experiments fall short of proving that that linguistic behavior is frequently reflexive rather than intentional, since the relevant intentions may be opaque even to their bearer, it does provide telling evidence to that effect.

Furthermore, these experiments explain why intentionalism is so intuitive. Subjects have the phenomenology of intention even when the subjective experience of intending does not match their actual behavior. Consequently, when we as philosophers or linguists reflect on our own linguistic behavior, our intuitions, guided by the phenomenology, will tend towards intentionalism. But since the phenomenology can be misleading, our intuitions are unreliable, undermining any armchair evidence in favor of intentionalism. If we really do utter without intending, as Lind et al. argue, intentionalism is in trouble.
III. Responses to TOMI

Denying the empirical inference

As the examples above demonstrate, multiple lines of empirical evidence seem to support TOMI. One avenue of response for the intentionalist is to produce an alternative explanation for the problematic empirical observations. The most plausible of these responses appeals to failures in executive function (EF), the mental capacity to regulate, control, and inhibit other cognitive functions. Some evidence (Bradford et al. 2015, Ferguson et al. 2015) seems to indicate that taking the perspective of others is particularly demanding on EF, perhaps because it requires inhibiting egocentric processes. If this is so, apparent theory of mind failures are consistent with the presence of theory-of-mind-laden processing, since we can attribute egocentric behavior to EF failure instead. Thompson (2014), for instance, defends intentionalism in the face of the false-belief task by attributing children’s failure to a deficit in EF, not theory of mind, citing Onishi and Baillargeon (2005) as evidence. The idea is that young children are aware of Sally’s false belief, but lack the mental control to keep themselves from sharing their own knowledge instead. Similarly, the intense world hypothesis (Markram, Rinaldi, and Markram 2007), denies that autistic individuals lack theory of mind, and gives a similar EF explanation for why they fail the false-belief task. Following the lead of these responses, we could argue that Keysar’s experimental results are due not to absence of intentional reasoning, but instead to failure to inhibit egocentric cognitive processes. If this line of response is correct, then higher order intentions might always be present in linguistic communication, but they sometimes fail to win out in competition with other cognitive determinants of behavior.

These alternative explanations for evidence purporting to undermine intentionalism are worth pursuing, but the orthodox interpretations of the false-belief experiment, Keysar’s work, and so on tell against intentionalism. Even if the alternative explanation turns out to be correct in some cases, it seems unlikely that this will be so in every case. Additionally, even if children, individuals with
autism, and so forth do have some theory of mind\textsuperscript{11}, it seems incontrovertible that they are unable to exercise it fully. This line of response is thus predicated on a shaky empirical bet.

Moreover, even if intentional reasoning is always present but sometimes suppressed, mere presence is not enough to guarantee the truth of intentionalism. Intentionalism is meant as an explanation of linguistic communication, and in cases where theory of mind is suppressed, the actual explanation for the communicative behavior will rest on the egocentric processes which won out. Arguments from failures of executive function thus fail to undermine TOMI, and the intentionalist must look for a stronger line of response.

\textit{Tweaking the intentionalist criteria}

Since dismissing the empirical data is unlikely to succeed, a defender of intentionalism might have to bite the bullet, granting the empirical cases. This bullet-biting can take two forms: we can weaken the intentionalist criteria so as to recover the previously excluded cases, or we can allow that many cases of linguistic communication fail to meet our criteria. An example of the first route is Moore (2013), who argues that Lewisian conventions can get off the ground even if we substitute ostention and imitation for some of the higher-order mental representations required in Lewis' original theory. Moves like Moore's successfully escape the empirical objection, but I doubt that we can really characterize them as satisfactory responses for the intentionalist. Yes, we can take Lewisian or neo-Gricean theories and swap out the intentionalist criteria for substitutes such as imitation and joint attention, but the resulting theories are not intentionalist at all, because they no longer implicate theory of mind in linguistic communication. The explanatory power of intentionalism comes from the sophisticated reasoning it attributes to linguistic agents, and much of this power is lost if we remove this sophistication. To bite the bullet in this way is to give up the game.

\textsuperscript{11} Young children, even infants, do seem capable of attributing agency of some sort to external beings, as shown in experiments discussed in Carey (2009) ch. 5. The ability to exercise theory of mind, however, involves being capable of deploying attributions of intention in various sorts of reasoning, not mere recognition of intentional states. The false belief task and similar experiments seem to show that young children lack the ability to fully exercise theory of mind in this sense.
**Rational reconstruction**

A better option for the intentionalist is hold to the strict definition of intentionalism but relax its empirical scope. In one common form, this approach argues that intentionalism is not a claim of descriptive psychology, but instead a rational reconstruction. Rational reconstructions are not meant to be predictive, or to describe actual causal mechanisms, but instead are meant make sense of a phenomenon by interpreting it as rational activity. Philosophers of language sometimes argue that objections like TOMI miss the mark because intentionalism is merely a rational reconstruction.

“Semantic and pragmatic theories are rational reconstructions of the ability of speaker-hearers to interpret uses of sentences,” claims Soames, “The cognitive processes by which this occurs are not our concern” (2010: 171-72). Likewise Bach argues that “Grice did not intend his account of how implicatures are recognised as a psychological theory nor even as a cognitive model. He intended it as a rational reconstruction...He was not foolishly engaged in psychological speculation about the nature of or even the temporal sequence of the cognitive processes that implements that logic” (2006: 25, see also Saul 2002, O’Rourke 2003). For reasons I’ll discuss below, I’m not convinced that this is a fair representation of Grice. It is clear however, that some intentionalist philosophers can escape the force of TOMI because the aim of their theorizing—rational reconstruction—does not require empirical accuracy.

As a general response to TOMI, however, the retreat to rational reconstruction is inadequate. Even if it is true that intentionalist philosophers of language only intend to produce rational reconstructions\(^\text{12}\), many intentionalists are scientific psychologists and linguistics. Additionally, psychologists and linguists draw on the work of philosophers of language (especially Grice and Austin) in crafting scientific theories. Scientific theories, unlike rational reconstructions, generally aim for explanation in terms of actual causes\(^\text{13}\), in this case an account of the actual psychology

---

\(^{12}\) Some philosophers who are intentionalists clearly aim for more than rational reconstruction, but arguing over who falls into this category doesn’t concern us here.

\(^{13}\) We might also want to use rational reconstruction to produce an account of ordinary language users’ *post hoc* interpretations of their own language use, perhaps because such interpretations play a role in holding each other to account for “what is said” (see Camp 2006). This aim, however, is mostly distinct from explaining primary linguistic cognition.
underlying linguistic communication. Borg has articulated precisely this problem for rational reconstructions of linguistic meaning (2009: 34):

My worry is that talk of ‘rational reconstruction’ runs the risk of driving too great a wedge between the semantic theory and the psychological theory, for if all one is offering is a way in which speaker meaning could be recovered, with no requirement that ordinary speakers do recover meaning in this way, then we seem to be sliding away from a picture which treats semantic content as dependent on psychological content and towards an account which treats semantic content and psychological content as more or less independent of each other.

Empirically inclined intentionalists thus can’t evade TOMI by retreat to rational reconstruction. Nevertheless, the retreat to rational reconstruction prefigures a tactic that will work for intentionalists in general, one in the spirit, if not the letter of Grice.

**IV. Intentionalism as idealized modeling**

In *Aspects of Reason*, Grice characterizes his inquiry into reasoning as an act of creating “models or ideal constructions” which can help us “understand actual reasonings” (2001: 7-8). If we extend the scope of these remarks to Grice’s work on meaning, it gives us the beginning of a general response to TOMI. Grice takes his philosophizing to be model building in a broad sense, including the construction of “analytic models,” which are meant to aid in philosophical categorization, and “normative models,” which provide examples for agents to aspire to (2001: 8). Neither of these types of models will rescue intentionalism as a descriptive psychological theory. But Grice also takes himself to be putting forward “explanatory models” which “play a central part in providing” an explanation of the phenomenon in question (2001: 8). In the case of reasoning, for instance, a model of reasoning drawn from first order logic might not accurately describe normal human reasoning processes, but it does explain why the actual processes make successful inferences (2001:9-10). If we apply the same sort of treatment to intentionalism, this leads us to see intentionalist theories not as direct descriptions of the psychological processes necessarily involved in linguistic communication,

---

14 He makes similar, but less explicit, remarks about meaning in “Meaning Revisited” (Grice 1991).
but as *idealized scientific models* of communication.

Scientific models can be employed for a variety of purposes, including prediction, but even predictive models are not meant to perfectly represent or simulate their target systems (Weisberg 2013). Instead, when scientists create a model, "the real systems in their respective domains of inquiry are knowingly and systematically misrepresented" (Jones 2005: 2). These misrepresentations, or *idealizations*, serve a number of purposes, including making it easier to analyze models and highlighting particular attributes of systems of interest. Idealization occurs in every science. Physicists construct mathematical models which omit the effects of friction or the gravitational forces exerted by small objects; biologists sometimes ignore the fact that populations are composed of finite numbers of discrete individuals; economists sometimes pretend that consumers are better informed than they actually are. These intentional falsehoods make complex problems tractable, which is why models can useful even if they are not strictly true.

Such idealized models are central to scientific practice, and intentionalists can avoid the empirical objections by construing their positions as putting forward scientific models of communication. Like Grice, we can see intentionalism as putting forward an ideal case of communication. Instead of Grice’s *normatively* ideal case, however, scientific intentionalists focus on a *descriptively* ideal (idealized) case—a model. Because models are idealized, they are necessarily false. The mere fact that a model inaccurately describes a number of empirical cases does not, on its own, constitute a reason to reject the model. Whether a model is a good model is a pragmatic question: does the model improve our understanding or increase our predictive or explanatory abilities? An idealization is licensed if it serves these ends, and unwarranted if it does not. To show that we can save intentionalism by construing it as a modeling strategy, then, I need to demonstrate that the empirical exceptions to intentionalism are the product of justifiable idealizations. To establish this, I will outline how intentionalist models of communication work as *special case models* and *minimal models*.

*Special case models*
A number of scientific fields use models of special cases, such as optimal or limiting cases, to great success. Examples include evolutionary optimality modeling, rational choice theory, and ideal observer analysis. The idea behind this sort of modeling is that we can “describe one class of cases, which are simple and tractable, and use these as the basis for a more indirect understanding of the others” (Godfrey-Smith 2009: 4). Cases of communication which depend on higher-order communicative intentions are just such an illuminating special case in a number of ways.

First of all, for predictive purposes, that the actual psychology of communication does not always meet the intentionalist criteria is of secondary concern. As famously pointed out by Dennett, in several domains of science we find it helpful in making predictions to ascribe intentions even where we suspect none exist. This is true in biology when, in order to reconstruct evolutionary history, we pretend that organisms were designed with intent (Dennett 1996). And it is true in psychology, when we use folk-psychological ascriptions to help us predict others' behavior and to bootstrap our understanding of mental processes (Dennett 1989). So it should come as no surprise that intentionalist models can serve as useful predictors of linguistic behavior. Modeling the special case of intentional communication allows us to make predictions about linguistic behavior in general.

Of course, if our requirement for the application of the intentional stance is mere predictive success, it overgenerates. We can make successful predictions in almost any domain by ascribing intentions—we can pretend that positively charged particles desire to attract negatively charged particles, and thus make successful predictions about electromagnetism, for instance. But scientists do not find it useful to create predictive intentional models in particle physics. Why then, is the intentional stance useful in biology and psychology? Because, unlike the case of particle physics, the intentions we ascribe are motivated by a parallel between biological needs and the motivations of a hypothetical agent. Evolution acts as if it had intentions because the adaptive pressures driving evolution are the same adaptive pressures that an intelligent designer would engineer organisms to meet. Sub-rational psychological process act as though they worked on an intentional level because they are responding to the same organismal needs that intentional reasoning does. The intentions we
might ascribe to positively charged particles, on the other hand, are ad hoc. We can create predictively successful intentional models of particle physics only after we already understand the behavior of the relevant physical systems, because there is nothing in their behavior that responds to circumstances the way an intentional agent would. So application of the intentional stance for predictive purposes is licensed only when there is a motivated parallel between the system in question and how an intentional agent would act.

This criterion is clearly met in the case of intentional models of meaning. The communicative goals that linguistic organisms respond to—coordination, information transfer, deception, etc.—are equivalent whether or not the organism is employing theory of mind to meet those needs. Behavior in the special case, where higher-order communicative intentions are in play, will thus generally resemble behavior in other cases. For predictive purposes, then, the idealization to the special case of intentional communication is warranted. It will not, of course, yield perfect predictions, but neither do special case models in other domains. Such models remain in use because some predictive error is tolerable, particularly if offset by other factors such as ease of use and understandability.

As special case models, intentionalist models also contribute to explanation. To see how, consider how optimality models are used evolutionary biology. A typical case of optimality modeling in behavioral ecology runs as follows: we consider a number of factors contributing to an organism’s fitness, measure all but one, then, assuming that the organism is maximally fit given its environment, use the measured quantities to estimate the unknown quantity. For example, we might look at available food sources to estimate the average jaw size of a predatory species. Or we might estimate the danger per minute a bird exposes its offspring to in leaving the nest unattended in conjunction with a measure of how many calories it is able to gather per minute to infer the amount of time it spends foraging for food for its chicks.

As descriptive accounts of behavior and evolution, however, optimality models are open to Gould and Lewontin’s (1979; see also Orzack and Sober 1994) famous objection that these models falsely assume that organisms are always maximally adapted to their environment. Gould and
Lewontin's objection falls short for the same reason TOMI falls short. They fail to take into account the fact that optimality models are idealized by design, and can be evaluated only in respect to particular modeling goals. As Parker and Maynard Smith argue in response to Gould and Lewontin, optimality models "serve to improve our understanding about adaptations rather than to demonstrate that natural selection produces optimal solutions" (1990: 27; see also Potochnik 2009). Even when optimality models fail predictively, they help us understand which causal components of an evolutionary process contributed in any particular case—if optimality obtains, natural selection has almost certainly dominated, but if not, we need to look for other evolutionary forces as well.

Moreover, by providing a limiting case optimality models give us a baseline from which we can determine the degree of contribution of all sorts of evolutionary forces, not merely the adaptive ones. As a particularly tractable special case, optimality models sort out these contributing factors and thus contribute to scientific explanation.

Similar useful features explain the persistence of ideal-agent rational choice modeling long after researchers (Simon 1955) called attention to their empirical inadequacy. To give one example, when looking for rationality-approximating heuristics the rational choice model still gives researchers a baseline to compare with the heuristic's performance. See, for instance, the methodology used by Czerlinski, Gigerenzer, and Goldstein (1999), where they compare the predictive success of hypothesized psychological heuristics against the performance of statistically-sophisticated ideal rational agents. By showing that the Take the Best heuristic performs nearly as well as multiple regression models, they show why human psychology would employ the Take the Best: it performs nearly as well as is possible, but only requires limited resources. And by answering that 'why' question, they provide an explanation for the phenomenon. Human reasoners only employ multiple regression in decision making on rare occasions. Nevertheless, a model of an agent in this special case helps us understand why actual decision makers use the simple heuristics they generally

---

15 This is similar to the methodology used in ideal observer studies of perception, another case illustrating the same points.
do. Additionally, it sets a baseline for maximum possible success that helps us understand why failures of reasoning can be considered failures. Thus, this special case model of an ideal rational agent is explanatory in the same way that Grice takes his account of an ideal reasoner to be explanatory in *Aspects* (2001:9-10).

Intentionalism sometimes functions as an explanatory special case model in the same way as optimality and rational choice models. For example, consider Lee and Pinker’s (2010) game-theoretic analysis of indirect speech acts. They present a scenario in which a speeding motorist, pulled over by police, attempts to offer a bribe in place of a ticket. The explanandum is that the motorist offers the bribe indirectly, saying “So maybe the best thing to do would be to take care of that here” rather than something like “If I pay you $50, will you let me off with a warning?” Lee and Pinker give a mathematically-modeled explanation, which relies on both participants reasoning about the other’s communicative intentions. It is beyond unlikely that an actual motorist and officer in such a situation would actually do all the math in Lee and Pinker’s explanation. Conversely, it is rather likely that the motorist would not engage in higher-order reasoning about the officer’s higher order reasoning about the motorist’s communicative intention. This may be, for instance, because the motorist has made a habit of offering bribes with that precise phrase. Despite these empirical inadequacies, however, the game-theoretic model presents an explanatorily illuminating special case. Motorists offer indirect bribes for the same reason whether or not they are engaging in mathematical calculations and higher-order intentional reasoning: plausible deniability. Moreover, the special case of the game-theoretic, intentionalist reasoner explains by way of contrast why the motorist who offers a straightforward bribe has committed a communicative anomaly. As this example illustrates, by idealizing to a special case, intentionalism can produce predictive and explanatorily successful models.

*Minimal models*

Intentionalist models are also sometimes *minimal models*, models which abstract away from the complexity of actual systems to focus on “core causal factors which give rise to a phenomenon”
Consider supply-demand models of pricing. Market prices are the result of a number of factors, including information asymmetries, cognitive biases, government intervention, etc. The paradigmatic economic model idealizes away from all these and presents price as a function merely of supply and demand. Because supply and demand are frequently the core causal factors, this idealization produces a minimal model which attains explanatory and predictive success while maintaining simplicity and tractability.

Communication is messy. Sometimes higher-order intentions play an important role, but so do a number of other factors. Nevertheless, insofar as higher-order communicative intentions are frequently core causal factors in linguistic communication, intentionalism can yield successful minimal models of linguistic communication. Take, for example, Clark’s (1996) account of language use, which models interlocutors as cooperatively keeping track of mutual knowledge, a behavior requiring higher-order reasoning about mental states. Clark’s model is highly idealized. As Keysar’s experiments demonstrate, speakers fail to keep perfect track of mutual knowledge. Additionally, linguistic communication is embedded in a complex set of activities, and cooperation and the common ground of mutual knowledge are not the only factors in play. They are frequently, however, core causal factors—Clark presents compelling empirical evidence that interlocutors often make linguistic decisions on their basis. Abstracting away from other complicating features of communication thus yields a relatively tractably minimal model which Clark uses to explain diverse features of language use. His model, for instance, allows him to explain the essential conversational role of non-verbal signals of understanding, such as head nods and hand gestures, as a means for interlocutors to coordinate on what is mutual knowledge. Similarly, Clark explains why linguistic behavior differs when there are more than two interlocutors by appealing to the fact that speakers in such situations recognize that they share different mutual knowledge with different participants. By focusing on a single core feature of communication—the role of higher-order intentions—Clark thus produces an explanatorily powerful minimal model.

Minimal models such as these are meant to be used in tandem with methods which do not
make the same idealizations. This means that models of meaning can be productive when used in conjunction with other philosophical, psychological or neuroscientific models. On the account of intentionalism as minimal modeling developed here, one can without conflict embrace intentionalism as well as conventionalism, psychologism, or most other accounts of meaning, since they too idealize away from some features of communication to focus on a limited number of important factors. Thus, even though intentionalism is empirically inadequate, as a modeling strategy it can help us understand and explain the psychology of human communication.

V. The virtues of construing intentionalism as idealized modeling

I have argued that although the facts motivating TOMI are probably true, the objection does not undermine the value of intentionalism. Intentionalists provide idealized models of a sort similar to special case and minimal models in other disciplines. These models are strictly speaking false, but purposefully so. Their validity is determined by their usefulness, not their truth. I have aimed to show how intentionalist models of communication are useful, but the strongest evidence of their utility is their persistent ubiquity in research on linguistic communication.

Treating intentionalism as a modeling strategy is successful where other responses to TOMI fail for a number of reasons. It does not require us to place risky empirical bets that future experiment will overthrow widely-accepted counterexamples to theory of mind use in linguistic communication. Nor does it have to give up on the scope of intentionalism, since intentionalist models can be useful across all the domains intentionalists work on. Moreover, it accomplishes all this without being overly reductive. In treating intentionalism as a modeling strategy we do not take something as complex as linguistic meaning and reduce it to a single factor. Instead, we provide a tool for exploring one aspect of human communication; a tool, for that matter, which works well in conjunction with our other tools.

Perhaps the most important virtue of my response to the empirical objection is that it captures what intentionalists actually seem to be doing. They abstract and idealize away from some features and instances of linguistic cognition to focus on particularly illuminating factors and cases, and they
refine the resulting models, combine them with other models, use them to guide experimental
design—all key features of model-driven science. My account is even in the spirit of Grice. Recall
that in Aspects he sees himself as creating idealized models, so in construing intentionalism as
creating idealized models we not only save it from the strongest objection against it, but we recover
those Gricean insights.

References

limits on theory of mind use? Evidence from adults’ ability to follow instructions from an ignorant speaker. The
Quarterly Journal of Experimental Psychology, 63(6), 1201-1217.


subsequent real speaking. Proceedings of the National Academy of Sciences, 111(49), 17678-17683.


Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). "How good are simple heuristics?" In Simple
heuristics that make us smart. Oxford University Press.


Weisberg, M. (2013). *Simulation and similarity: using models to understand the world*. OUP.