AUTHENTIC SPEECH AND INSINCERITY*

People often use language to talk to themselves—reminding, encouraging, chastising—as well as to talk to others. This very innocent observation can, if allowed, ripple through the dogmas in current speech-act theory, crushing many long-standing, cherished ideas. Or so I will argue. Specifically, it turns out that a very influential theory of insincere speech and insincere directives seems plausible only when self-addressed speech is disregarded. But I can be insincere to myself as well as to others. Moreover, paying attention to self-talk, in this context, points the way to a new perspective on the authenticity of speakers, understood as the combination of sincerity and genuineness.

I start, in section I, by spelling out the standard account of directive insincerity. This is, roughly, the idea that speakers are insincere in directing someone to do something if that is not something they desire. Here I propose two distinctions which will be important later. First, I distinguish the attitude the speaker wants to produce in the addressee, by making the utterance, from the attitude the speaker has in the production of the utterance. Second, focusing on the attitude produced, this should be broken down into the attitude-part and the content-part. I propose that the attitude is one of intention and that the content is a prescription rather than a proposition; thus, if I tell you to jump, and I am genuine, what I want is to produce in you the intention to jump.

In sections II–IV I have two aims. First, employing the distinctions already introduced, I pinpoint the exact sense in which the standard account assumes an asymmetry between directives and assertives. Second, I argue that this asymmetry only appears plausible because theorists have failed to consider the case of self-addressed speech. I conclude that any theory, like the standard account, which implies this asymmetry is thereby cast in doubt.

Finally, in sections V–VII, I develop an alternative theory in some detail. As mentioned, this is based on the insight that there is a distinc-

tion between sincerity and authenticity in speech acts. Taking assertion as the paradigm case, speakers are insincere in asserting *p* if they do not believe *p*. But, moreover, the speech act is fake or not genuine if the speaker asserts *p* but does not want or intend the addressee, thereby, to believe *p*. The assertion is authentic, however, when the speaker both believes *p* and wants the addressee to believe *p*. This distinction is readily applied to directive utterances and self-addressed speech.

In this final part, I also show how this view supports the idea that so-called bald-face lies are not really lies, but rather inauthentic assertions. And I point out how the resulting theory of directive competence predicts that higher-order communicative intentions are involved in the maturation of pragmatic and linguistic competence from a very early stage of development.

I. THE STANDARD ACCOUNT

The standard account of insincere speech is often traced to John Searle, and its basic structure has been endorsed by many since.[1] Assertion is normally considered the paradigm case.

INSINCERE ASSERTIVE

S is insincere in asserting *p* if and only if *S* does not believe *p*.

Many philosophers have argued that even this simple account needs to be revised. Most of these proposals have to do with the role of the mental state of belief. Some have argued, for example, that the belief must be conscious or that it must be higher order. But otherwise the basic idea has retained its currency.[2] As is well known, however, Searle also proposed sincerity conditions for other types of speech acts. And it is commonly assumed in the literature that sincerity ought to apply to more than only assertive utterances.[3]

Here I focus on directives. Typically, directive speech acts are understood as utterances whereby the speaker directs the addressee to

---

[1] John Searle, *Speech Acts: An Essay in the Philosophy of Language* (Cambridge, UK: Cambridge University Press, 1969).

[2] Timothy Chan and Guy Kahane, "The Trouble with Being Sincere," *Canadian Journal of Philosophy*, XLI, 2 (2011): 215–34; John Eriksson, "Straight Talk: Conceptions of Sincerity in Speech," *Philosophical Studies*, CLIII, 2 (2011): 213–34; Richard Moran, "Problems of Sincerity," *Proceedings of the Aristotelian Society*, CV, 3 (2005): 341–61; Michael Ridge, "Sincerity and Expressivism," *Philosophical Studies*, CXXXI, 2 (2006): 487–510; Jeremy Schwartz, "Saying 'Thank You' and Meaning It," *Australasian Journal of Philosophy*, CXCVIII, 4 (2020): 718–31; and Andreas Stokke, "Insincerity," *Noûs*, XLVIII, 3 (2014): 496–520.

[3] David Kaplan, "The Meaning of 'Ouch' and 'Oops'," *Howison Lecture in Philosophy*, 2004, University of California, Berkeley; and Stokke, "Insincerity," *op. cit.*

perform some action.[4] The standard account of directive insincerity, from Searle, is the following.[5]

INSINCERE DIRECTIVE

S is insincere in directing H to v if and only if S does not want H to v.

This certainly strikes a chord in common sense, and roughly the same idea was proposed earlier by R. M. Hare.[6] Hare would say that S is insincere if S does not *intend H* to v. For my purposes, it will not be necessary to distinguish between different pro-attitudes—desire, preference, intention—in formulating the standard account, as the argument will apply to every theory of this general kind. Thus, the speaker is directively insincere, on the standard view, if and only if they lack some favored pro-attitude toward H's v-ing.

To take a simple illustration, assume that I say to you,

(1) (You) go home now!

In uttering (1), I will normally have performed the directive speech act of requesting or commanding that you go home now. In English the second-person pronoun is optional. But if I utter (1) without actually wanting or intending you to go home it seems like something has gone wrong. On the standard account, I have failed to satisfy the sincerity condition for directives; to be sincere I must have a specific type of attitude, namely the desire or intention that the addressee perform the action I have requested. This pro-attitude appears to be part of what the speaker communicates by performing the directive act.[7] At least, the speaker can reasonably predict that, on the basis of hearing (1), the addressee can infer that the speaker probably *wants* the addressee to go home.

But how exactly could the speaker fail to have the relevant pro-attitude in uttering (1)? A number of theorists have argued, in similar fashion, that this is impossible and, so, directives cannot be insincere, or are automatically sincere.[8] My argument against the standard ac-

[4] Daniel Boisvert and Kirk Ludwig, "Semantics for Non-declaratives," in Ernest Lepore and Barry Smith, eds., *The Oxford Handbook of Philosophy of Language* (Oxford: Oxford University Press, 2006), pp. 864–92.

[5] Searle, *Speech Acts, op. cit.*, p. 66.

[6] R. M. Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952), p. 13.

[7] Cleo Condoravdi and Sven Lauer, "Imperatives: Meaning and Illocutionary Force," *Empirical Issues in Syntax and Semantics*, IX (2012): 37–58, at p. 38.

[8] Condoravdi and Lauer, "Imperatives," *op. cit.*, p. 43; Jane Heal, "Insincerity and Commands," *Proceedings of the Aristotelian Society*, LXXVII (1977): 183–201; and Mark Siebel, "What Is an Illocutionary Point?," in Günther Grewendorf and Georg Meggle, eds., *Speech Acts, Mind, and Social Reality: Discussions with John R. Searle* (Dordrecht, the Netherlands: Springer, 2002), pp. 125–39, at p. 129.

count does not depend on this kind of impossibility and so, for sake of argument, I will assume that directives can be issued by speakers who lack the relevant pro-attitude. On the positive theory developed here, these will be fake or non-genuine rather than insincere, unless they are *self-directed*, in which case they could turn out to be insincere as well (see section v).

As Andreas Stokke has argued, the Searlean standard account can be generalized by saying, roughly, that insincerity consists in the communication of an attitude which the speaker lacks.[9] I think this is a fruitful idea and will take it on board here as a working hypothesis. For example, the generalization will cover declarative and non-declarative utterances alike. To assert p insincerely is to communicate the belief that p while lacking that belief. To direct H to v insincerely is to communicate some pro-attitude toward H's v-ing while lacking that attitude. Of course, the notion of what the speaker communicates is very broad, as a single utterance might communicate a number of different attitudes, and Stokke proposes that we restrict ourselves to consciously held attitudes. My argument exploits this feature to show, in effect, that the standard account fails to respect a structural symmetry between assertives and directives, which can be motivated independently. Roughly, if we make sure to separate the different contents conveyed by a directive utterance, we will alight upon a contender in the determination of insincerity. And this contender turns out to unearth the commonalities between insincere assertion and insincere directives, which have remained hidden to theorists so far.

But which contents, more precisely, are communicated by a literal directive? In the terminology of Condoravdi and Lauer, there is a distinction between what the directive *expresses* and what it *conveys*.[10] On their view, (1) would express the content, roughly: *the addressee goes home now.* But it also conveys *that the speaker wants the addressee to go home now.* Condoravdi and Lauer follow the standard view in thinking that the latter would be relevant for insincerity, concluding, however, that the speaker's desire is always present and thus directives are automatically sincere.[11] But I think the nature of the content expressed deserves more careful attention in this debate. As we shall see, it is important to acknowledge the role of the kind of attitude the addressee is supposed to bear to the expressed content.

---

[9] Stokke, "Insincerity," *op. cit.*
[10] Condoravdi and Lauer, "Imperatives," *op. cit.*, p. 38.
[11] *Ibid.*, p. 43.

At least since G. E. M. Anscombe, it has been common to separate the expression of propositions from the expression of *prescriptions*.[12] At a first pass, this can be understood as a distinction between two groups of *attitudes* speakers can bear to one and the same representation. The first is the group of *co-attitudes*, including belief and knowledge, which are supposed to represent the world truly or accurately. Any co-attitude that *p* aims to co-exist or otherwise correlate with the truth of the representation. The second is the aforementioned group of *pro-attitudes*, including desire and intention, whose satisfaction consists in *bringing about* the truth of the representation. In Searle's influential terminology, beliefs would have a word-to-world direction of fit and intentions a world-to-word direction of fit.[13] Take "milk," for example, which could either belong to an inventory or a shopping list. In the first case, let us say, the representation is true if and only if there is milk in the fridge. In the second, it guides someone's action to make it so. To simplify, we can suppose that in both cases "milk" expresses the same representation—something like: *that there is milk in the fridge*—and that the difference consists in the fact that thinkers can stand in different attitude relations to that representation. I can either believe or intend to make it true that there is milk in the fridge. I will drop this simplification shortly, however. Finally, we could then say that writing "milk" on a list is either intended to activate the *p*-belief or the *p*-intention in the addressee, depending on whether it is an inventory or a shopping list. Directives seem, thus, to involve the expression of prescriptions and pro-attitudes.

To see this more vividly, consider three interesting features of directive acts. First, when I utter (1) to direct you to go home, I will not have said anything true or false. This can be tested by noting the infelicity of responding with "That's true/false." Second, it follows that I cannot have asserted anything, at least not directly, if assertion is understood as a speech act whose expressed content can be believed to be true or false. Certainly, there are theories according to which (1) expresses a modalized proposition, something along the lines of: you *should* go home now.[14] But one way to understand such a view is to distinguish, again, between what is expressed and what is conveyed. Thus, even if the latter involves a modalized proposition, there

---

[12] G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957), section 32. See, for example, Peter B. M. Vranas, "New Foundations for Imperative Logic I: Logical Connectives, Consistency, and Quantifiers," *Noûs*, XLII, 4 (2008): 529–72.

[13] John Searle, *Expression and Meaning: Studies in the Theory of Speech Acts* (Cambridge, UK: Cambridge University Press, 1979), pp. 3–4.

[14] Magdalena Kaufmann, *Interpreting Imperatives* (Dordrecht, the Netherlands: Springer, 2012).

will be some level of expression at which the content is robustly non-truth-conditional, as indicated by true/false tests. As Craige Roberts notes, we should prefer theories that can capture the non-assertive and non-truth-conditional nature of (1), if at all possible.[15] Third, the expressed content of a directive is, in some sense, essentially indexed or anchored in the addressee. The directive can only be "satisfied" if the addressee performs some action compatible with what the speaker communicates. For example, (1) is not satisfied by some audience member other than the addressee going home now.[16]

To respect these differences, we require something more fine-grained than the distinction between the content expressed and the content conveyed by a directive. A more detailed account will be spelled out in section v below, but the proposal is inspired by a particular Gricean approach to distinguishing assertives from directives.[17] To start, we need to distinguish the attitude *produced* from the attitude of *production*. The normal function of a directive utterance is to (attempt to) produce—induce, activate, generate, whatever—an *intention* in the addressee. This is the produced attitude. The attitude of production is, rather, the speaker's *intention to produce* the produced attitude. What I have been calling the content expressed is roughly equivalent to the produced attitude, while the content conveyed is roughly equivalent to the attitude of production. I go along with Jane Heal, and others, in holding that the speaker's aim is to produce the *attitude* conducive to compliance, rather than merely the act of complying.[18] This simple point calls for the structure of attitudes of production embedding produced attitudes.

---

[15] Craige Roberts, "Speech Acts in Discourse Context," in Daniel Fogal, Daniel W. Harris, and Matt Moss, eds., *New Work on Speech Acts* (Oxford: Oxford University Press, 2018), pp. 317–59, at p. 331.

[16] According to Zanuttini, Pak, and Portner, the imperative clause is part of the more general 'jussive' clause type, which also includes promissives and exhortatives (these are all syntactic categories, not speech-act theoretic). Promissives have the speaker as subject while exhortatives are addressed to both speaker and hearer as a group. My argument only concerns 'literal directives', understood as directive speech acts where what the speaker utters is an imperative clause (Raffaella Zanuttini, Miok Pak, and Paul Portner, "A Syntactic Analysis of Interpretive Restrictions on Imperative, Promissive, and Exhortative Subjects," *Natural Language and Linguistic Theory*, xxx, 4 (2011): 1231–74).

[17] Paul Grice, "Utterer's Meaning, Sentence-Meaning, and Word-Meaning," *Foundations of Language*, iv (1968): 225–42; Paul Grice, "Utterer's Meaning and Intention," *The Philosophical Review*, lxxviii, 2 (1969): 147–77; Stephen Schiffer, *Meaning* (Oxford: Clarendon Press, 1972); Kent Bach and Robert Harnish, *Linguistic Communication and Speech Acts* (Cambridge, MA: MIT Press, 1979); and Daniel W. Harris, "Imperative Inference and Practical Rationality," *Philosophical Studies*, clxxix (2022): 1065–90.

[18] Heal, "Insincerity and Commands," *op. cit.*, p. 191.

We can see, already, how this picture suggests a strong analogy between directives and assertives: the normal function of the latter is to produce *beliefs*—or other co-attitudes—in exactly the same sense. Although this is not the focus of the current paper, and my arguments can play with other pro-attitudes, I will tend to talk in terms of the production of *intention* in the case of directives.[19]

Finally, happily, this helps to further analyze the cognitive effect produced by directives. Assume that 'I' stands for the attitude of intention and 'π' for the *prescriptions* embedded by the intention. On this picture, when I say (1) to you and mean what I say literally, what I *mean* or *express* is some mental state of the form $I(\pi)$. That is to say, I mean for you to form the intention to go home now.[20] We can also say that π is what I mean (prescribe, express), but that is merely a less informative specification, since I must also mean some attitude relation rather than another. I do not mean that you are supposed to *fear* or *hope* something, for example.[21] In the terminology adopted from Condoravdi and Lauer above, $I(\pi)$ is the form of the content expressed (or produced), while one part of what is *conveyed* will be the speaker's intention that the addressee form the $I(\pi)$-attitude, this being the attitude of production. In the positive account developed in section v below I will call the latter the speaker's *effective* intention.

This framework fits nicely with the three features articulated above. Intuitively, π is an action type (*to go home now*) and $I(\pi)$ is an intention to perform a token action of that type. In formal semantics we can, as is currently popular, think of the denotation of the imperative clause type as a property indexed or applied to the addressee. For example, Nate Charlow thinks of the denotation as a property of a *plan*, namely the property the plan has just if it requires that you go home now.[22] A plan is a set of intentions, obeying basic principles of practical rationality. If you adopt a plan with this particular property, which

---

[19] Harris argues specifically for this view, on the basis that otherwise we could not explain the infelicity of directing addressees to do something they lack control over. To take his example, I cannot felicitously command you to stop having nightmares, because you cannot form the intention to stop, at least not directly ("Imperative Inference and Practical Rationality," *op. cit.*, pp. 1069–70). Portner, Condoravdi and Lauer, and Starr would rather talk in terms of influencing (or expressing) desires or preferences (Paul Portner, "The Semantics of Imperatives within a Theory of Clause Types," in Robert B. Young, ed., *Proceedings of SALT XIV* (Ithaca, NY: Cornell University, 2004), pp. 235–52; Condoravdi and Lauer, "Imperatives," *op. cit.*; and William B. Starr, "A Preference Semantics for Imperatives," *Semantics and Pragmatics*, XIII (2020): 1–60).

[20] Heal, "Insincerity and Commands," *op. cit.*, p. 191.

[21] See, for example, Paul Grice, *The Conception of Value* (Oxford: Clarendon Press, 1991), pp. 63–64.

[22] Nate Charlow, "Logic and Semantics for Imperatives," *Journal of Philosophical Logic*, XLIII, 4 (2014): 617–64.

might be the upshot of someone saying (1) to you, that part of your plan is satisfied if and only if you go home now. Among other things, this would help to explain why (1) and "Don't go home now!" seem logically inconsistent, without being either true or false. Roughly, to adopt a plan with both properties is inconsistent in virtue of violating basic requirements of rationality.[23] The property denoted by the clause partly determines the action type intended by the speaker in issuing the directive utterance. Anyway, neither the action type nor the intention seem to have intuitive truth conditions, and the formation of an $I(\pi)$-state—such as the intention to go home now—is more like a decision to act than a judgment about something being true. And this is exactly what a directive utterance, on this view, is supposed to achieve: the addressee's decision to form the attitude the speaker intends to produce.

II. WHY THE ASYMMETRY?

In this section I will argue that the standard account of directive insincerity introduces an implausible asymmetry between sincere directives and sincere assertives. Later, I will explain why the asymmetry is implausible, but it arises, I believe, because theorists tend to consider only *other-directed* speech in their accounts. This comes as no surprise, as many have assumed that the primary function—in some sense—of language is interpersonal information exchange. Even so, speakers engage in episodes of self-addressed private speech all the time, for various communicative or non-communicative purposes.[24] The current argument does not require any specific assumptions about the nature or function of inner or private speech, only the obvious fact that it exists.

Following this line of thinking, let us call speech acts 'paradigm' if they are primarily directed at an addressee other than the speaker. Now, there is one very striking difference between paradigm as-

---

[23] See also Nate Charlow, "Clause-Type, Force, and Normative Judgment in the Semantics of Imperatives," in Fogal, Harris, and Moss, eds., *New Work on Speech Acts*, *op. cit.*, pp. 67–98; Harris, "Imperative Inference," *op. cit.*; Paul Portner, "Imperatives and Modals," *Natural Language Semantics*, XV, 4 (2007): 351–83; Paul Portner, *Mood* (Oxford: Oxford University Press, 2018); and Roberts, "Speech Acts in Discourse Context," *op. cit.*

[24] See, for example, Daniel Gregory, "Inner Speech, Imagined Speech, and Auditory Verbal Hallucinations," *Review of Philosophy and Psychology*, VII, 3 (2016): 653–73; Peter Langland-Hassan and Agustin Vicente, eds., *Inner Speech: New Voices* (Oxford: Oxford University Press, 2018); Adam Winsler, Charles Fernyhough, and Ignacio Montero, eds., *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation* (Cambridge, UK: Cambridge University Press, 2009); and Elmar Unnsteinsson, "The Social Epistemology of Introspection," *Mind and Language*, XXXVIII, 3 (2023): 925–42.

sertives and paradigm directives. According to the working hypothesis adapted from Stokke, the former will be sincere in virtue of the speaker's own private relation to the produced attitude. That is to say, if I want to produce the belief that p, I am sincere if and only if I myself *believe* that p. But a parallel definition appears incoherent in the latter case, because the produced attitude is not a potential state of the speaker at all. What I mean or express in uttering (1) is, if our assumptions so far are right, an intention to go home now, and the intention is supposed to be formed by you, the addressee. We can rummage around in my mind all we want, but we will never find there an intention to go home now which will be relevant to the determination of whether or not my utterance is made sincerely. For example, my utterance does not become sincere if *I* intend to go home now. And I cannot even form *your* intention to go home now, because your intention has the function of guiding and controlling *your* conduct in a manner in which my intentions cannot. Certainly, there is an action type both of us can intend to perform, namely to go home now. But our intentions in performing the same action type are still essentially different in that my intentions guide my actions and not yours, and vice versa.

As indicated before, this difference generalizes to any pro-attitude we might want to use in theorizing about directives. Take preferences, for example, so that the attitude produced will be: *your preference to go home now.* Again, there is no attitude in my mind which matches or is identical to your preference to go home now, such that my attitude will plausibly determine the utterance's sincerity conditions. I might prefer to go home now, but that is irrelevant. I might prefer that you go home now, but that is not the attitude produced (it is not even an attitude of yours). Of course, the option remains to say, in response, that sincerity is determined by the attitude of production and not by the attitude produced. But my reply will have to come later, after I have argued that structurally analogous speech acts will have structurally analogous properties (section IV).

Anyway, it follows that the insincerity of a directive cannot be determined by a direct mismatch between the $I(\pi)$-attitude produced and the producer's own $I(\pi)$-attitude. That is to say, I am not insincere in virtue of lacking the pro-attitude I aim primarily to communicate and produce. Rather, on the standard account, I am insincere in virtue of lacking some second-order attitude which embeds the produced attitude: that I *want* you to form the $I(\pi)$-attitude (this would be the attitude of production). This is the asymmetry between insincere assertives and insincere directives on the standard account. Because assertions are supposed to produce belief in propositions—or so we will

assume here—rather than intentions to act, they are very different. My belief that p and your belief that p are both true if and only if p, so they are essentially the same even if realized by different people. That is to say, their satisfaction conditions, or truth conditions, are identical. What I assert is something I myself can either believe or fail to believe, and so my own first-order relation to the proposition expressed may, plausibly, determine my insincerity.

To drive the point home, notice that the asymmetry survives a comparison between addressee-indexed assertions and directives. So, the asymmetry does not derive from the fact that, normally, directive content mandates indexing, while for assertives this is optional.

(2) You went home yesterday.

By assumption, if I say (2) to you, I speak insincerely if and only if I do not believe that you went home yesterday. No one would be inclined to think that my insincerity in uttering (2) is due to the fact that I do not really want you to believe that you went home yesterday. This indicates that the apparent asymmetry does not arise because the content expressed is somehow indexed to the addressee.

Jane Heal and Mark Siebel argue that various other asymmetries between assertives and directives indicate that there really is no such thing as an insincere directive.[25] Condoravdi and Lauer propose that sincerity is automatic for directives.[26] By contrast, the standard account assigns blame to a co-conspirator. Since the real offender is not available, we get the next best thing, namely, the speaker's second-order desire or intention to produce the intention to π in the addressee. This is the attitude of production, which always embeds the attitude produced. And, as already mentioned, this may seem like the right conclusion. If I order you to go home when I really want you not to go home, I seem—if this is possible—to be speaking insincerely or dishonestly. But why exactly should we accept this particular asymmetry? On reflection, as I shall now go on to argue, it seems unmotivated.

III. THE ARGUMENT FROM SELF-DIRECTED SPEECH

Now, let us consider utterances that fail to adhere to the paradigm of other-directedness. Specifically, I argue, it turns out that *self-directed* directives can be sincere or insincere just like assertives and for the same reasons. If I plan to go home at 3pm and, when I see that it is

---

[25] Heal, "Insincerity and Commands," *op. cit.*; and Siebel, "What Is an Illocutionary Point?," *op. cit.*

[26] Condoravdi and Lauer, "Imperatives," *op. cit.*

now exactly 3pm, I say (1) to myself, I am being sincere to myself, and literally so. The content of what I say or mean matches my own intentional state, and I utter (1), plausibly, to cause the activation or further operation of that state, for example, in working memory. To write "milk" on the shopping list can also be used as an example here. If the list writer is directively sincere, they in fact intend to buy milk and, presumably, want that intention to be re-activated at some later time. If the shopping list is exclusively addressed to someone else—here I ignore the interesting case of directives which are simultaneously self- and other-addressed, sometimes called 'exhortatives' (see footnote 16)—sincerity of this sort is impossible, for reasons already canvassed. On the other hand, if the writer has no intention at all to buy milk, neither at the time of writing nor at the expected time of reading, the utterance was literally insincere.

More precisely, if π is the directive *content*, the prescription expressed, and $I(\pi)$ is the attitude I want to produce by my utterance, there can be a perfect match between the attitude produced and my first-order mental state. I utter (1) to produce the intention to π, while simultaneously having an intention to π. The attitude of production is the more complex intention to produce the intention to π. To make this even more obvious, perhaps, imagine that I write (1) in a self-addressed email which is timed to appear in my inbox at 3pm. Even as an interpreter, I must recognize my own prior intention to produce an intention to go home, in myself. Here, sincerity and insincerity work in exactly the same manner as in paradigm, other-directed assertives.

Imagine, to take a different example, that someone appears at the edge of jumping out of an airplane with a parachute on, but really has no intention of going through with it. It seems coherent to suppose that they utter "Jump!," wanting thereby to produce an intention to jump in themselves, even if they simply do not intend to jump, and never intended to jump. For this to be rational, the speaker only needs to believe that it is *possible* that the utterance will be effective in producing the intention, not that it is all that likely. Perhaps the example is more plausible when the speaker knows that others can hear these self-directed commands—the display is also supposed to inform others of something—but that does not affect the main point. The point is that directives can be insincere with respect to what the speaker says or expresses, rather than merely with respect to some second-order desire embedding the $I(\pi)$-attitude meant by the speaker. Addiction, weakness of will, and wishful thinking can also provide familiar examples. A smoker could have a first-order desire to smoke and a higher-order desire not to have that desire. A self-

directed utterance of "Stop!" might thus be intended literally to activate the intention not to pick up the cigarette.

Why exactly is this a problem for the standard account? Well, it seems like the account only allows for two easy reactions, both dead-ends. First, one could say that there is no such thing as an insincere directive which is self-addressed or self-directed in the requisite sense. But this flies in the face of common sense and a great deal of empirical evidence. Inner or private speech is a well-studied phenomenon, and directives are certainly thought to play a significant role, for example, in sport psychology.[27] If so, we should predict that insincerity may feature as part of that role, just as it does when we lie to ourselves in self-directed assertions.

Second, and more promisingly, proponents of the standard account could insist, still, that self-directed insincerity consists in the speaker's lacking the relevant second-order desire. We would need an explanation of why this is not true also in the case of assertion; that is, why is my self-addressed assertion then not insincere in virtue of me not wanting myself to believe what I assert? Even so, the idea might be to question the very distinction between second-order and first-order states in self-directed speech acts. What exactly is the difference, one might ask, between S wanting S to (intend to) go home now and S wanting or intending to go home now?

To examine this distinction in self-directed insincerity we need an example where the second-order desire is present while the speaker lacks the first-order desire or intention. And this was in fact the structure of the skydiver's case. There is no intention to jump, but still there could be a desire to produce the intentional act of jumping by way of the self-directed command. The idea was to say that this is an insincere directive, because the *sincere* case would be one where the skydiver really intends to jump all along and makes the utterance as a reminder or piece of self-encouragement. Thus there would be a perfect match between the speaker's attitude in making the utterance (*intention to jump*) and the attitude produced or communicated (again, *intention to jump*). The presence of the second-order desire—the attitude of production—does not make the utterance sincere, if the difference between the first- and second-order states is a real one. We can compare this to assertion for illustration. Even if I want myself to believe that I am not bald, saying to myself in the mirror "I'm not bald," I am still lying to myself if I really believe that I am bald. I conclude that the standard account implies an asymmetry between

---

[27] See James Hardy, "Speaking Clearly: A Critical Review of the Self-Talk Literature," *Psychology of Sport and Exercise*, VII, 1 (2006): 81–97.

insincere directives and insincere assertions. But some might argue that this is not a real problem.

<div align="center">IV. IS THE ASYMMETRY IMPLAUSIBLE?</div>

Why should we think that the asymmetry is implausible?[28] It is certainly not incoherent. Basically, the default assumption should be that the cognitive mechanism of pragmatic competence is uniform across the capacity to produce assertions and directives. On the Gricean view, to be explained in more detail later, pragmatic competence is the capacity to perform speech acts with a set of specific audience-directed intentions. Three clause types are normally considered universal or near-universal in human languages: declarative, imperative, and interrogative.[29] Data from language acquisition suggests that the first two are acquired at the same or similar stages of development.[30] This supports the assumption that the capacity to perform the corresponding speech acts is explained by one and the same cognitive mechanism. Besides, positing a single structure for both capacities—insincere assertives and insincere directives—is more parsimonious.

On this view, we already posit the relevant mechanism to explain assertive competence, by which I mean that even very early in the development of pragmatic competence, speakers can perform normal assertive acts, and that this is due to some mechanism in virtue of which the relevant intention types can be formed and employed in the control of planned action. My claim here is that the default assumption should be that directive competence recruits exactly the same resources. If so, we should expect the attitude of production to have a certain structure. The speaker has (1) a pro-attitude toward producing a particular (2) pro-attitude in the addressee. If this is right, we should expect speech acts to be insincere in virtue of a relation between (2) and the speaker's own pro-attitudes. But this only seems possible in self-directed speech, and this point has not been noted in the literature to date. The current proposal has the virtue of identifying a parameter which immediately explains the asymmetry that remains, which is precisely the distinction between self and other.

---

[28] I would like to thank a reviewer for this JOURNAL for raising the question, which gave rise to this section.

[29] Jerrold M. Sadock and Arnold M. Zwicky, "Speech Act Distinctions in Syntax," in Timothy Shopin, ed., *Language Typology and Syntactic Description I: Clause Structure* (Cambridge, UK: Cambridge University Press, 1985), pp. 155–96; and Roberts, "Speech Acts in Discourse Context," *op. cit.*

[30] Alexandra Aikhenvald, *Imperatives and Commands* (Oxford: Oxford University Press, 2010), pp. 325–30.

One might wonder, still, how this could be so. Will we not still predict the same asymmetry for all sorts of other speech acts, like questions, promises, and apologies? If so, it would seem like we should not be too bothered by it in the case of directives.

In my view I could have replaced directives with questions and made many of the points I have made so far. If we accept recent arguments from Whitcomb, Friedman, and Carruthers, we should think of interrogative attitudes as 'basic' like intentions and beliefs.[31] Further, interrogative clauses seem to be a human linguistic universal specifically suited for the expression of this attitude. But there are two important differences. First, interrogative attitudes are more controversial than intentions. Second, interrogative attitudes are easily shared like assertive (that is, belief-like) attitudes. If it is my intention to produce in you curiosity about whether *p*, by uttering an interrogative, I might be insincere in virtue of not being curious myself about whether *p*. Now, I am not denying that intentions can be shared in some sense, only that a shared intention is what is directly meant or expressed by uttering a literal directive.

Promises and apologies are different because they are more likely to be relative to factors like culture, convention, and history, which probably has effects on the cognitive mechanisms involved and the process of acquisition. But we do make promises and apologies to ourselves, which is certainly relevant and interesting. It is equally clear that apology, for example, is not a conventional device for the production of a distinctive attitude, different from intention or belief. To simplify, assume that apology is associated with the attitude of regret. Still, apology is not in the business of producing regret; more plausibly, it produces the *belief* that the speaker feels regret. Consider a simplified account:[32]

Apology

S apologizes for *v*-ing by uttering *X* if *S* thereby intends the addressee to believe that *S* regrets having *v*-ed.

Note that this allows for the idea that insincere apologies are apologies. And if this is right, the Gricean theorist is free to posit additional structure into the attitude embedded by the effective intention (the attitude of production), and this seems to be a very satisfying way to

capture the difference between apology and directives. Most importantly, this fits the schema I have adapted from Stokke; speech acts are insincere in virtue of the speaker lacking the produced attitude. I take this as further reason to accept the schema as a general theory of insincerity across different types of speech acts.

There are many reasons, however, not to focus on highly conventionalized speech acts like apology or promising, only one of which is how they can vary between groups of speakers. The most pressing reason is the added complexity in the accompanying notion of insincerity. Is the apology insincere in virtue of the speaker's lack of regret, lack of belief that they are regretful, or both? Does the belief, the regret, or both need to be consciously accessible? The same questions will arise on a similarly Gricean story about promises:[33]

PROMISE

S promises to *v* by uttering *X* if *S* thereby intends the addressee to believe that uttering *X* makes it obligatory that *S* intend to *v*.

Although I think something *like* this must be right, I will not argue for it here. I only want to point out that the theory retains a clear distinction between the attitude of production (that is, effective intention) and the produced attitude, and that this allows for a general explanation of insincerity which is grounded in a theory of more basic speech-act types. The insincere promise may consist in *S*'s lacking the belief that the utterance puts *S* under any obligation. But it seems like the promise could also be insincere in virtue of *S* lacking the intention to *v*. That is not a big problem, because in that case lacking the intention might be part of the explanation why *S* lacks the belief in question. Note also that these accounts of apology and promises adhere to specific Grice-inspired methodological strictures.[34] That is to say, they are *speaker-centric*; promises and apologies are insincere in virtue of something about the speaker—and what the speaker means—not in virtue of something about the addressee's rational or normal reaction. That being said, promises and apologies are sometimes embedded in highly institutionalized and conventionalized practices which can have significant consequences for our taxonomy.

---

[33] Again, very much simplified from Bach and Harnish, *ibid.*, p. 50.

[34] See, for example, Kent Bach, "Context *ex Machina*," in Zoltán Gendler Szabó, ed., *Semantics versus Pragmatics* (Oxford: Clarendon Press, 2005), pp. 15–44; Jerry Fodor and Ernie Lepore, "Out of Context," *Proceedings and Addresses of the American Philosophical Association*, LXXVIII, 2 (2004): 77–94; and Stephen Neale, "Silent Reference," in Gary Ostertag, ed., *Meanings and Other Things: Themes from the Work of Stephen Schiffer* (Oxford: Oxford University Press, 2016), pp. 229–344. I will turn to this matter again in section VI.

More to our purpose, it seems like the potential generality of this Gricean account adds greatly to its plausibility. That is to say, we seem justified in thinking that insincere speech, generally, consists in a mismatch between the produced attitude and the speaker's own attitudes. Since there is also good reason to suppose that literal directives are conventional devices for producing intentions, just like literal assertives are supposed to produce beliefs, this general structure appears to fail in the case of insincere directives. Further, if what I have argued in this section is right, we can conclude that this structural disanalogy or asymmetry, implied by the standard account, is implausible. We are justified in looking elsewhere. But what is the alternative?

V. AUTHENTICITY AND SINCERITY

What we need is a distinction between sincere speech and *authentic* speech. To start, distinguish two intentions speakers have in performing communicative speech acts, according to the broadly Gricean framework mentioned above. First, there is the *effective* intention, which is the intention to have some specific cognitive effect on the addressee. This is what we have, so far, been calling the attitude of production. It is roughly equivalent to the second-order desire in the standard account of directives, applied to all communicative acts. For example, if I tell H that p by uttering something I will normally intend H to form the belief that p or, at least, to form some specific co-attitude A with respect to the proposition that p. The cognitive effect intended by the speaker, then, is going to be some $A(p)$-state formed by the addressee, in response to perceiving the utterance.

Second, on this framework, speakers will also have a *signaling* intention, which is the intention to get the hearer to recognize the effective intention. As Scott-Phillips would put it, the speaker wants to signal the signalhood of the utterance.[35] The addressee's recognition of the signaling intention is then considered sufficient for *communicative* success.[36] When addressees satisfy the speaker's effective intention—by actually forming the attitude the speaker intends to produce—the speaker is more than merely communicatively successful. Combina-

[35] Thom Scott-Phillips, *Speaking Our Minds: Why Human Communication Is Different, and How Language Evolved to Make It Special* (New York: Palgrave Macmillan, 2015).

[36] For example, Paul Grice, *Studies in the Way of Words* (Cambridge, MA: Harvard University Press, 1989); Stephen Neale, "Paul Grice and the Philosophy of Language," *Linguistics and Philosophy*, xv, 5 (1992): 509–59; Dan Sperber and Deirdre Wilson, *Relevance: Communication and Cognition* (Oxford: Blackwell, 1986); Schiffer, *Meaning, op. cit.*; and Mandy Simons, "Local Pragmatics in a Gricean Framework," *Inquiry*, lx, 5 (2017): 466–92.

tions of effective and signaling intentions can be called *communicative intentions.*[37]

To generalize this theory, my act of trying to communicate $A(p)$ will be sincere if and only if I bear the attitude A to p. The same schema will apply to directives—with the qualification about the difference between self and other, to be discussed below—by replacing the cognitive effect with $I(\pi)$. Furthermore, my $A(p)$-assertion, directed at addressee H, is *genuine* with respect to $A(p)$ if and only if I really intend to produce the $A(p)$-state in H. On this view, genuineness is determined by the effective intention, and my communicative act is *fake*, with respect to $A(p)$, if the corresponding effective intention is lacking. Finally, we can say that the assertion is *authentic* if and only if it is both sincere and genuine. It is inauthentic, then, exactly when it is both insincere and fake, with respect to one and the same cognitive effect.

An utterance can be fake but sincere with respect to the same attitude, for example, in sincere irony.[38] Consider Grice's example of uttering (3), ironically or sarcastically, to mean that someone is the opposite of a good friend.[39]

(3) He's a fine friend.

On his theory, the speaker merely *makes as if* to say that Harald, to give him a name, is a fine friend—so, does not really *say* this—and thereby conveys that something else is really meant. What the speaker means is often thought to be, in some intuitive sense, the opposite of what was made as if to say.[40] Mandy Simons would say, in this case, that p is what the speaker merely *expresses*—without saying or meaning— and Stephen Neale describes the phenomenon in terms of *play-saying*

---

[37] Here I am only slightly modifying the standard terminology in relevance theory, where 'effective' is 'informative' and 'signaling' is 'communicative' (Sperber and Wilson, *Relevance, op. cit.*; and Robyn Carston, *Thoughts and Utterances: The Pragmatics of Explicit Communication* (Oxford: Blackwell, 2002)). I think 'effective' is better here because directives are not obviously a matter of information exchange.

[38] See Andreas Stokke, *Lying and Insincerity* (Oxford: Oxford University Press, 2018), p. 174.

[39] Grice, *Studies in the Way of Words, op. cit.*, p. 34.

[40] I will not argue for this theory of irony or sarcasm here, but for a helpful discussion, see Elisabeth Camp, "Sarcasm, Pretense, and the Semantics/Pragmatics Distinction," *Noûs*, XLVI, 4 (2012): 587–634. For other theories of irony, see, for example, Gregory Currie, "Why Irony Is Pretence," in Shaun Nichols, ed., *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction* (Oxford: Clarendon Press, 2006), pp. 111–33; Ernie Lepore and Matthew Stone, *Imagination and Convention: Distinguishing Grammar and Inference in Language* (Oxford: Oxford University Press, 2015); and Deirdre Wilson and Dan Sperber, *Meaning and Relevance* (Cambridge, UK: Cambridge University Press, 2012), chapter 6.

and *play-meaning* that p.[41] Let us assume, then, that what the ironical speaker play-asserts by uttering (3) is a cognitive effect of the form $A(p)$, namely the belief that the man is a fine friend. Now, either the speaker believes p or does not believe p. In standard cases of irony, speakers do not believe the proposition they play-assert; instead, they only believe what they really mean (that he is a jerk). But this is not essential to irony on the Gricean view.

That is to say, I could utter (3) ironically to communicate my belief that Harald is an awful friend, while believing that he is a fine friend (the $A(p)$-state). If so, my utterance is *fake* but *sincere* with respect to what is play-asserted, namely $A(p)$. The utterance is fake in this manner because I lack the effective intention to produce, by my utterance, the $A(p)$-state in the addressee. But it is sincere with respect to this very state because it matches my own attitude. Importantly, it does not follow that I have no effective intention at all; the effective intention just involves a different cognitive effect. If this is right, irony can be fake and sincere with respect to what the speaker makes as if to say. Actors who happen to believe what they say on stage could also be engaged in fake but sincere assertions, in virtue of lacking the corresponding effective intentions. When actor A says to B, on stage, "I am your mother," she does not want or intend to activate in B the belief that A is B's mother. It is only part of the script. But if, by some historical accident, the belief is true and A is desperately—given the unlucky circumstances—trying to hide it from B, the utterance is sincere but fake with respect to what A play-asserts.

Normally, however, irony is inauthentic with respect to one and the same content, for the speaker neither believes nor intends to produce the belief that, in this example, Harald is a fine friend. Rather, the speaker intends the hearer to recognize that $A(p)$ is *not* meant, so some other cognitive effect can be inferred as what *is* meant (that Harald is a jerk).

These distinctions are readily applied to directives. Directives can be sincere or insincere only when self-directed, as already argued. But other-directed directives are *fake*, exactly like assertives, when the speaker lacks the relevant effective intention. This is the case, for example, in directive irony. To borrow an example from Aikhenvald, imagine that your partner comes home late one night and, in a drunken rampage, starts trashing the place.[42] You say,

(4) That's right. Break everything in the house!

---

[41] Simons, "Local Pragmatics in a Gricean Framework," *op. cit.*; and Neale, "Silent Reference," *op. cit.*, p. 258.

[42] Aikhenvald, *Imperatives and Commands*, *op. cit.*, p. 234.

There is a cognitive effect here which you merely express or give voice to, namely your partner's intention to break everything in the house. If uttered ironically, (4) is really intended to produce a very different effect in your partner: the opposite effect, it may seem.[43] Thus, the directive is fake with respect to this merely expressed or play-directed effect. For concreteness, let us assume that there are two cognitive effects at issue:

(5) *I*(*break!*)

(6) *I*(*don't break!*)

On this view, by uttering (4), your partner fakes (5) in order to convey a different cognitive effect, namely (6). As an act of communication, then, the utterance is successful if and only if your partner recognizes your effective intention, namely that they should not wilfully break everything in the house (6).

Of course, this theory of irony is controversial, but it serves as a convenient illustration (see footnote 40). The distinction between fake and genuine speech acts, understood in terms of effective intentions, applies much more widely. Consider, briefly, directives with foregone conclusions. If I am tied to a chair while some hoodlums rummage through my stuff, I could say, "Take the money! What do I care!" The directive definitely lacks the associated effective intention and is thereby fake or non-genuine. That is to say, I faked a directive with the content, roughly: *intention to take the money*. Here the speaker's further communicative purpose is less transparent, but the utterance is fake rather than insincere with respect to this particular content. It is not insincere, as already argued, because the attitude is not available to me as a first-order mental state. It is important to notice that speech acts are defined, here, as fake, insincere, or inauthentic only with respect to some specified cognitive effect or other. Faked effects are in the service of the effects actually intended by the speaker.

So far, I have avoided the topic of *lying*, which has received a lot of attention recently.[44] In the literature on lies, theorists have become interested in the notion of *bald-faced* lies, which seems relevant here.

---

[43] Grice's account of implicature would require some slight modification to accommodate the notion that speakers can implicate something by a directive, if directive contents lack truth conditions. The *Maxim of Quality* is formulated in terms of the speaker's attention to *truth* and *evidence*. For one model of how to do this, see David Braun's account of how speakers can implicate by asking questions. Part of his proposal is to replace the *Maxim of Quality* with a more general *Maxim of Sincerity* (David Braun, "Implicating Questions," *Mind and Language*, XXVI, 5 (2011): 574–95).

[44] Thomas L. Carson, *Lying and Deception: Theory and Practice* (Oxford: Oxford University Press, 2010); Eliot Michaelson and Andreas Stokke, eds., *Lying: Language, Knowledge,*

Lies are bald-faced, I take it, when the speaker does not disguise the fact the utterance is a lie. Many argue, on this basis, that lies do not require the speaker to have an intention to deceive the addressee. It is common ground that the speaker does not believe what is said, so deception does not seem to be part of the speaker's goal in making the utterance.[45]

My argument so far would suggest, instead, that bald-faced lies are assertions that are both fake and insincere. So, relative to a specific cognitive effect, $A(p)$, the speaker lacks the intention to produce that effect in the addressee and, at the same time, does not bear the attitude A to the proposition p. This agrees, more or less, with the views of a group of philosophers who argue that bald-faced lies are not really assertions at all, essentially on the basis that the relevant effective intention, or something very much like it, is missing.[46] On the theory proposed here, bald-faced lies are cases where the speaker's expression of a particular attitude is not authentic. So, for example, if a defendant takes the stand and swears to the jury that they did not commit a crime, they might be speaking inauthentically, in part because everyone already knows that they are guilty. Even so, the defendant might not want to be on the record admitting their guilt. But if the truth is already common ground, it seems like the speaker neither believes p nor intends the audience to form the belief that p on the basis of the utterance. Thus, some other intention or purpose must be guiding the speech act.

Finally, we can now see clearly how this theory can unify the notions of sincerity and authenticity across speech-act types. The differences only show up in highly predictable situations; there is no such thing as other-addressed directive insincerity, only non-genuineness or faking. Faking is understood, both for assertives and directives, in terms of the speaker lacking the effective intention with respect to a par-

---

Ethics, and Politics (Oxford: Oxford University Press, 2018); Jennifer M. Saul, *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics* (Oxford: Oxford University Press, 2012); and Stokke, *Lying and Insincerity, op. cit.*

[45] Thomas L. Carson, "The Definition of Lying," *Noûs*, XL, 2 (2006): 284–306; Don Fallis, "Lying as a Violation of Grice's First Maxim of Quality," *Dialectica*, LXVI, 4 (2012): 563–81; Saul, *Lying, Misleading, and What Is Said, op. cit.*; Roy Sorensen, "Bald-Faced Lies! Lying without the Intent to Deceive," *Pacific Philosophical Quarterly*, LXXXVIII, 2 (2007): 251–64; and Stokke, *Lying and Insincerity, op. cit.*

[46] Daniel W. Harris, "Intentionalism and Bald-Faced Lies," *Inquiry* (2020), https://doi.org/10.1080/0020174X.2020.1775381; Jessica Keiser, "Bald-Faced Lies: How to Make a Move in a Language Game without Making a Move in a Conversation," *Philosophical Studies*, CLXXIII, 2 (2016): 461–77; and Ishani Maitra, "Lying, Acting, and Asserting," in Michaelson and Stokke, eds., *Lying: Language, Knowledge, Ethics, and Politics, op. cit.*, pp. 65–82.

ticular attitude. Directive sincerity still exists, but it is only defined for self-directed speech. I have not yet given any examples of directive speech acts that fail to be *authentic*. They are easy to come by, however. Consider the timid skydiver again, and call him Marshall. He says (7) to himself, but this time he does so only in order to placate the others watching. He wants everyone to think that he is seriously trying to overcome his fear, but has long since decided that he will sooner die than jump from an airplane.

(7) Jump!

To explain Marshall's overall plan, it seems that we must posit a specific cognitive effect which is merely faked or play-meant, something like (8).

(8) *the intention to jump*

On this picture, he utters (7) to himself, and part of his goal is to fake (8) as the effect or attitude he thereby means to activate or produce in himself. Marshall's real purpose, however, is to make the others believe that he really intends to jump, or intends to incite himself to jump. These others are what Herbert Clark would call 'side participants' in the conversation.[47] Marshall is both the speaker and an addressee, but (7) is uttered in part to inform side participants that he has performed a speech act intended to produce the attitude in (8) in himself. Of course, the side participants are not supposed to recognize that (8) is the *faked* effect of uttering (7). Indeed, Marshall knowingly lacks the effective intention of producing (8) in himself by uttering (7) and does not possess the attitude involved. Thus, the directive act is inauthentic with respect to (8). Plausibly, the utterance is genuine but insincere with respect to some cognitive effect Marshall is trying to generate in the minds of side participants. Here is a possible candidate:

(9) *the belief that Marshall intends to jump*

That is to say, Marshall may utter "Jump!" to produce (9) in the others, while (9) is not one of his actual beliefs. If so, the utterance is insincere but not fake with respect to (9).

This theory of authentic speech constitutes a serious and well-motivated alternative to the standard Searlean account of directive

---

[47] Herbert H. Clark, *Arenas of Language Use* (Chicago: University of Chicago Press, 1992), p. 202.

insincerity. The theory is based on an empirically plausible and fruit-ful model of human communicative capacities—so-called pragmatic competence—namely, the broadly Gricean or intentionalist program in pragmatics. I have already argued that the standard account fails to respect plausible symmetries between assertives and directives. In what follows, I address one objection and provide an additional rea-son to prefer the Gricean view to the Searlean.

VI. TRUSTING THE SPEAKER

Despite what I have argued, one might wonder why exactly it matters that the attitude produced by a directive does not match any of the speaker's own attitudes. More specifically, in placing our trust in the speaker, by deciding to comply with what we are directed to do, it seems like we rely on their intention that we do something. And in-sincerity, here, appears to be connected to the potential for misplaced trust; the speaker turns out to lack the attitude we were relying on.[48]

As mentioned briefly before (section IV), I think the perspective of the addressee can obscure important distinctions. It is quite true that the addressee relies on the speaker in the way suggested, but still it is important to distinguish the roles of different attitudes in ground-ing trust. Also, insincerity is a property of the speaker's act, regardless of how the addressee reacts. Focus on assertion first. When I assert p to you sincerely and truthfully, I have both the effective intention of producing an $A(p)$-state in you and the signaling intention that you recognize my effective intention. In forming the $A(p)$-state you corre-spondingly rely on two things: first, the fact that I am in the $A(p)$-state myself (I am sincere), and second, the fact that I genuinely intend to produce the $A(p)$-state in you (I am genuine). I have proposed to call this an authentic speech act.

So, I might express $A(p)$ ironically, without intending that you thereby form that belief. Even if I do believe p, you will not or should not rely on me in forming your belief that p. This is because I was not genuine in expressing $A(p)$. Thus, your trust in me is of a dif-ferent sort in these cases, and we should not lump these two notions together under the heading of 'insincerity'. That is to say, I am insin-cere when I do not believe what I assert, but I am doing something different when I do not intend you to believe what I make as if to assert.

Now we can come back to directives. It is true that we rely on the speaker's attitude of production in forming our own intention. This

---

[48] Thanks to a reviewer for this JOURNAL for raising the questions relevant to this section.

means that we are relying on the speaker's effective intention, which embeds a certain other intention. So, on my view, we are relying on the speaker's genuineness: we form our own intention in light of our confidence that the speaker genuinely wants us to form that intention. Yet, because self-talk is a real phenomenon which our theories should account for, there will be two types of reliance for directives too.

Consider the example of the meeting once more. If I realize suddenly that it is now 3pm, and that my meeting was supposed to start at 3pm, I might say to myself: *Go now*! In an 'authentic' case, I will rely on my own intention as well as my intention to produce that intention. To make the possibility of failure more obvious, let us introduce delay, like before. I send myself an email which is supposed to show up in my inbox at 3pm. It says: *Go to the meeting now*!

There is a sense in which I will rely on my own intention to go to the meeting. Thus I rely on the attitude produced rather than the attitude of production. If I have no intention to go to the meeting in making the utterance—writing the email—my reliance as addressee is spurious to that extent. This holds whether or not I did have the intention to produce the intention to go to the meeting. Further, the spuriousness in question is well described as insincerity. In this case I may successfully fool my later self into thinking that my earlier self intended to go to the meeting, and this is what makes the utterance insincere. It is not made sincere, holding everything else fixed, by assuming that my earlier self was *genuine*. That is, just like in the case of assertion, speakers are not sincere merely by having the effective intention. And, certainly, introducing the delay is optional; the same distinctions will apply in principle without it.

We are considering a case where my later self trusts my earlier self to have really had the intention to go to the meeting. That is to say, my later self forms the intention to go, partly on the basis of the belief that my earlier self had this first-order intention. If so, my later self's reliance parallels exactly the reliance of another on my asserted belief. Thus my later trust is spurious in virtue of being based on insincerity but not in virtue of being based on faking, as these terms are understood here. Admittedly, this becomes even more complicated if we include the so-called "third clause" (sometimes known as the "Gricean" or "directive" intention) with the effective and signaling intentions.[49] For this clause says that the speaker intends the addressee

---

[49] For discussion, see Manuel García-Carpintero, "Gricean Rational Reconstruction and the Semantics/Pragmatics Distinction," *Synthese*, CXXVIII, 1–2 (2001): 93–131, at

to form the relevant attitude *partly on the basis* of recognizing the effective intention. In fact I think the third clause should be included and, so, that the complications are welcome.[50] For current purposes, however, we can assume that the third clause is optional. I conclude that the relevant distinctions between how different attitudes ground differences in the addressee's reliance are available for both assertives and directives. And, to see that this is so, all we need to do is to take self-talk seriously.

VII. SOME PREDICTIONS

Assuming that the Gricean theorist is correct to insist that there are two pro-attitudes involved in directives—roughly, the attitude produced and the attitude of production—it can be shown that the standard account is implausible. Basically, we are owed an explanation why one of these attitudes is chosen as the one expressed rather than the other, and the choice will seem arbitrary. On the standard account, apparently, the attitude of production is chosen because it is indeed one which is available to the speaker as a first-order attitude. The speaker, in their own mind, forms the effective intention itself. But this will not do as a reason; everyone already accepts that speakers can express and communicate attitudes which they do not have, for example, in paradigm insincere assertion, expressing the belief that *p* while not believing that *p*. So, we are owed an account of why the speaker expresses the second-order attitude, embedding the addressee's intention, rather than the first-order attitude itself (the produced attitude). Assertion is simpler because the attitude expressed and the attitude produced may be identical, it is assumed, so the choice seems unnecessary. But, as we have seen, if we take self-talk seriously, we must also take seriously the idea that speakers can sometimes express *their own* first-order intentions, in issuing directives, as well as any second-order desire for the addressee's compliance.

On the Gricean framework, directive competence is the capacity to utter imperative clauses with a complex communicative intention which embeds the addressee's intention to act. The corresponding interpretive competence, for directives, is the capacity to recognize speakers' communicative intentions when they properly embed prescriptions rather than propositions. The paradigm is a directive, like

pp. 102–03; Neale, "Paul Grice and the Philosophy of Language," *op. cit.*, pp. 547–49; and Dan Sperber and Deirdre Wilson, "Beyond Speaker's Meaning," *Croatian Journal of Philosophy*, xv, 44 (2015): 117–49.
  [50] See Elmar Unnsteinsson, *Talking About: An Intentionalist Theory of Reference* (Oxford: Oxford University Press, 2022), chapter 6.

(1), addressed to someone else, with a literal meaning of the form $I(\pi)$. This identifies the produced attitude with what the speaker literally means. Even in such a standardized use of the directive, what the speaker literally means cannot correspond to any of their first-order mental states. This state is not available to the speaker, consciously or unconsciously. That is to say, the speaker cannot form, in their own mind, the mental state which, occurring in the mind of the addressee, constitutes the satisfaction of the meaning of the directive utterance, if the meaning is an attitude like $I(\pi)$. Thus the speaker must stand in some *other* relation to that attitude, presumably a higher-order or metarepresentational relation.

It follows that any competent user of imperatives as directives, addressing others, must distinguish between the attitude they *mean* and their own attitude in producing the utterance. In the Gricean terminology introduced here, they must distinguish between the state they intend to produce and their own mental state in performing an act whereby they attempt to produce that state. And this does not only apply to 'insincere' directives, but also paradigm ones. In the case of assertion, by comparison, some philosophers have held that sincerity is the default setting and that insincere speech—requiring some sensitivity to the distinction between what is asserted and what is believed—is best understood as pretended assertion.[51] On this view, effective intentions could, for all we know, be a late development in a speaker's pragmatic competence. No such move is available for directives, because even paradigm or 'sincere' utterances require the speaker's cognitive sensitivity to the very same distinction between produced attitudes and attitudes of production.

This indicates that directive competence is not acquired by being in some first-order state and then becoming able to express or give voice to that state directly, for the state is not available to the speaker as a first-order state. Speakers, even from a very young age, must then be sensitive to the distinction between the state they intend to produce and any state of their own, potentially long before they acquire the

---

[51] Elizabeth Fricker, "Against Gullibility," in Bimal Krishna Matilal and Arindam Chakrabarti, eds., *Knowing from Words* (Dordrecht, the Netherlands: Springer, 1994), pp. 125–61; Ruth Garrett Millikan, *Beyond Concepts: Unicepts, Language, and Natural Information* (Oxford: Oxford University Press, 2017); Richard Moran, "Problems of Sincerity," *Proceedings of the Aristotelian Society*, cv, 3 (2005): 341–61; Wilfrid Sellars, "Language as Thought and as Communication," *Philosophy and Phenomenological Research*, xxix, 4 (1969): 506–27; Bernard Williams, *Truth and Truthfulness: An Essay in Genealogy* (Princeton, NJ: Princeton University Press, 2002); and David M. Rosenthal, "Postscript to 'Intentionality'," in Stuart Silvers, ed., *Rerepresentations: Readings in the Philosophy of Mental Representation* (Dordrecht, the Netherlands: Kluwer, 1989), pp. 341–44.

competence to issue insincere assertions. To acquire this latter competence, however, is simply to apply the same distinction to a different kind of speech act, namely assertives. So, plausibly, if there is a period in the acquisition of pragmatic competence where speakers cannot issue insincere assertions, only sincere ones, it is because the effective intention, for some reason, embeds a state of the speaker when what is meant is a proposition rather than a prescription during that period. The reason is not that the speakers lack effective intentions, because they are already needed to explain directive competence. It is reasonable to suppose, then, that effective intentions should be generalized to cover both assertive and directive competence, for it is more likely that the same cognitive mechanism is responsible for both, rather than two distinct mechanisms. By the same token, the mechanism for effective intentions should be postulated for both paradigm and non-paradigm utterances.

As mentioned before, research in language acquisition suggests that children can acquire the imperative clause type very early, possibly earlier than other clause types.[52] If effective intentions are necessary for the acquisition of directive competence, as I have tried to argue, it follows that children can form effective intentions before acquiring the capacity to issue insincere assertions. It would then follow, moreover, that children have the requisite capacity for metarepresentation before or during language acquisition, as predicted by Gricean models.[53] The prediction has long been controversial but appears to be on sound footing. Basically, directives require speakers to engage in sophisticated reversals in perspective which may appear unnecessary in the competent use of assertives, at least until we reach insincere assertives. To satisfy a directive the competent addressee must be able to activate a specific mental state, while recognizing that it is not a state of the speaker, except insofar as its activation in the addressee is what the speaker wants. To use a directive competently, however, one must be able to reverse these two perspectives. Paying attention to directives and authenticity in speech seems to suggest that effective intentions are necessary for the development of pragmatic competence. The standard account of directive insincerity is, instead, expressivist; the idea is to find a first-order attitude which the speaker can express directly. As we have seen, however, the expressivist perspective fails to do justice to insincerity in self-directed speech.

---

[52] Aikhenvald, *Imperatives and Commands, op. cit.*, pp. 325–30.
[53] Scott-Phillips, *Speaking Our Minds, op. cit.*

VIII. CONCLUSION

If I sincerely remind myself that I should go now to the meeting I had planned at 3pm, because 3pm is now, it seems like I must intend to attend the meeting. That is to say, my intention to attend seems to determine the extent of my sincerity in speaking. This can be gauged, I have argued, by considering which changes would make my utterance insincere. For example, I might have no intention to attend. I might even have a policy of never attending meetings scheduled at 3pm. If, in that situation, I say to myself, "Go now!," it seems like I am insincere to myself. My reason for making the utterance could be to make some eavesdropper believe that I am getting ready to go, while I am not.

If this is right, what I have called the 'produced' attitude can sometimes determine the sincerity status of a directive utterance, namely when the attitude and the utterance are self-directed. This is a good reason, I believe, to accept the Grice-inspired distinction between attitudes of production and attitudes produced and, furthermore, to apply this distinction to any type of speech act. Moreover, by the failure of the standard account of insincerity, we are motivated to look for an alternative. I have argued that there is such an alternative view, which relativizes sincerity and related properties—authenticity, genuineness—to particular cognitive effects and to the distinction between self and other.

ELMAR UNNSTEINSSON
University College Dublin and University of Iceland