

# On the Uselessness of the Distinction between Ideal and Non-Ideal Theory (at least in the Philosophy of Language)

Herman Cappelen and Josh Dever

Forthcoming in *Routledge Companion to Social and Political Philosophy of Language* (eds Justin Khoo and Rachel Sterken)

There's an interesting debate in moral and political philosophy about the nature of, and relationship between, ideal and non-ideal theory. In this paper we discuss whether an analogous distinction can be drawn in philosophy of language. Our conclusion is negative: Even if you think that distinction can be put to work within moral and political philosophy, there's no useful way to extend it to work that has been done in the philosophy of language. Here is the plan:

- In section 1 we present a sketch of how theorizing in philosophy of language works, so that we can highlight the ways in which that theorizing can go wrong.
- In section 2, we outline four ways in which ideal and non-ideal theory have been understood in moral and political philosophy and ask whether any of them can provide frameworks to help us pick out interesting subsets of work in philosophy of language. We conclude that none of them can.
- In section 3 we turn to the question of whether the expression 'social and political philosophy of language' (the title of this volume) can be used to pick out an interesting sub-discipline. Again, we conclude that it cannot.
- In sections 4 we give a diagnosis of why the efforts in sections 2 and 3 fail and draw some general lessons from those failures.

Before we get to all that, some context to make clear why this an interesting issue. Some might have in mind the following little caricature of philosophy of language:

The philosopher of language sits in his armchair, never venturing into the outside world. He thinks of a few sentences, and then starts throwing mathematical machinery at those sentences. The result is a theory that handles some tiny fragment of the language

(probably mostly mathematical language), giving a bloodless picture of how some hypothetical perfectly rational, dispassionate, and cooperative agents might use that language to pass back and forth packets of information.

Confronted with this caricature, one might think that philosophy of language has lost track of “the real world”. Where is the language of metaphor, hyperbole, propaganda, and threats? Where are the liars, the bullshitters, the hucksters, and the rabble-rousers? Someone familiar with the debate in moral and political philosophy might think the right way to frame this concern is by appeal to the ideal/non-ideal distinction: Have philosophers of language been too focused on ideal philosophy of language and ignored non-ideal philosophy language?

We’re going to answer that question in the negative. In the end, we think that philosophy of language properly develops theoretical tools by starting with the study of simple cases which in fact make up the bulk of ordinary language use, and then systematically extending those theoretical tools to cover a wider and wider range of harder and more complicated cases that often involve navigating tricky waters where questions about specifically linguistic phenomena start to merge into broader questions about overall human behavior.

That’s not to say, of course, that philosophy of language never goes wrong. It does, as all theoretical projects do. But we will argue that attempts to import a dispute from normative theory about the relationship between ideal and non-ideal theory don’t find a distinctive failing in philosophy of language to latch onto. There are, for example, many underexplored topics in philosophy of language. Philosophers of language have had relatively little to say about the persuasive functioning of language in advertisement, and they have had relatively little to say about the semantics of prepositions, for example.<sup>1</sup> But there is no evidence of any *systematic patterns* in what topics are underexplored. It’s hard, of course, to get any good count of stuff we haven’t been paying attention to—but to the extent that looking around is a helpful guide, unexplored topics seem to be scattered about at random.

## 1. Theory Building: A Partial Sketch

We begin by observing one thing that goes under the name “idealization” that will *not* yield a helpful distinction between ideal and non-ideal theory. Consider a toy sketch of how building a theory of language goes. We begin by collecting some data. Data collection is a messy matter. If we just go out and start writing down what’s happening as people use language, we’re going to get an enormous collection of heterogeneous information. We’ll have details about people’s linguistic output; people’s intentions, plans, hopes, fears, desires, and anxieties; people’s professions, social status, clothing, financial difficulties, and mathematical proficiency; the social conventions governing politeness, greetings and departings, meals, and gift-giving; the

---

<sup>1</sup> Others have had much to say about both of these topics. People have looked, for example, at how syntactic complexity affects viewers’ comprehension and recall of advertisements (Bradley and Meeds (2002)), and there is work about prepositions by semanticists such as Joost Zwarts (e.g. 2017) .

specifics of the physical environment including the weather, the local geography, and what objects are positioned where.

No theory of language can incorporate all of that data—if it tried, it wouldn't be a *theory*, but just a data dump. The theorist must extract from the data what look to be interesting patterns, and then try to explain those patterns. Doing this involves both deciding what data to keep and what data to discard and deciding what categories to use in characterizing the patterns we find in the data. All of this can go under the heading of *idealizing*, although it could just as well be called *abstracting* or *simplifying*.<sup>2</sup>

But idealization in this sense is not going to produce a distinction between ideal and non-ideal theories. That's because *all* theory involves idealization in this sense, so there is no possibility of a "non-ideal" theory. To be non-ideal would be not to idealize, and thus not to abstract and extract patterns from the data. That wouldn't be a theory at all.

However, this sense of idealization does have something to do with our opening caricature of a philosophy of language that has lost engagement with the real world. When we abstract patterns from the data, we make choices about what to pay theoretical attention to. For example:

- We have to decide where to get our data (what parts of the total possible data to take into account). We could look at data restricted to our own introspective judgments on sentence types, or we could look at cross-linguistic field data drawn from language users in a variety of social settings.
- We have to decide what aspects of the data to try to explain. We could focus on the parts of the data concerning truth conditions of and inferential relations among sentences, or we could focus on the parts of the data that concern the forms of emotional, social, and financial manipulation that people are attempting in their language use.
- We have to decide what kinds of patterns to extract from the data. We could characterize patterns using categories like *rational and cooperative speaker*, thus perhaps finding a simpler pattern that applies in fewer cases, or we could characterize patterns using categories like *deceptive and uncooperative speaker*, thus perhaps finding a more complicated pattern that applies in more cases.

These choices about how to idealize/abstract/simplify our data aren't choices that will make the difference between *right* and *wrong* theories. Whichever way we choose to idealize, we will (assuming we then do our theory building well) end up with a correct theory. But the choices will make a difference in what our theory ends up being *about*. Idealize in one way, and we end

---

<sup>2</sup> Anticipating some of our discussion of the next section, this form of idealization maps reasonably well onto what Mills calls building ideal-as-descriptive models, although we will suggest that Mills's further distinction between ideal-as-descriptive and ideal-as-idealization models is not helpful.

up with a theory of cross-sentential anaphora. Idealize in another way, and we end up with a theory of the role of presuppositions in creating plausible linguistic deniability.

So while we don't get, on this picture of idealization, a distinction between ideal and non-ideal theory, we do get a distinction between explored and unexplored topics. How we idealize shapes what topics we end up theorizing about, and the result is that some topics might end up being neglected.

Concerns that important topics are being neglected are legitimate concerns. Work in philosophy of language inevitably sometimes doesn't gather data that it ought to gather (thus the increased interest in experimental methods and cross-linguistic work, for example). It inevitably sometimes neglects interesting patterns in the data that would have led to interesting projects (the curious behavior of 'or' in free choice permission claims was neglected during decades of work on the semantics of logical vocabulary). And it inevitably sometimes provides explanations that restrict the scope of theories unnecessarily (we start by building theories that focus on the semantics of declarative sentences, when it turns out that equally robust theories could have incorporated the semantic features of interrogative and imperative sentences as well).

None of this is meant to be startling news. We don't always get the right data, ask the right questions, or use the right tools in answering the questions we do ask. *Hopefully* "don't always" falls reasonably far short of "frequently don't", and hopefully there are mechanisms in the field that help fix the problems when they arise. (Note that the specific instances cited in the previous paragraph have more-or-less-satisfactorily been addressed.) But we should always have our eye on what topics have been unjustly neglected in the field.

## 2. 'Ideal vs Non-Ideal Theory'

We turn now to the various ways in which the terms 'ideal' and 'non-ideal' have been used in describing theories in moral and political philosophy. In these fields, there is a lively debate about how to balance ideal vs non-ideal theory, with many theorists arguing that the alleged dominance of ideal theory has had various pernicious consequences. The debates internal to moral and political philosophy will not be our concern here. Instead, we will focus on the question of whether this distinction and the accompanying debate can be applied to philosophy of language and communication. Is there a distinction between ideal and non-ideal philosophy of language? If yes, has ideal theory dominated? And has that had negative consequences?

To answer these questions, we first need to understand the distinction as it has been drawn in moral and political philosophy. As with most kinds of fashionable jargon, we find that the terminology is used in many conflicting ways. That in itself is a serious problem (and a strong *prima facie* reason for avoiding the spread of the terminology to other fields). In what follows we outline some of the uses of these terms in moral and political philosophy and the reasons why they don't illuminate anything useful about work done in philosophy of language. The

conclusion will be this: no version of the ideal/non-ideal distinction terminology helps delineate anything important about linguistic theorizing. There are no good reasons to let that piece of confused terminology spread to philosophy of language.

Before we turn to the use of '[non]ideal theory' in moral and political philosophy, let's reemphasize one central point made in the first section and give it a name:

**Underexplored:** In any field, there will always be topics that are under-explored. However, to say that you don't need to import terminology from a contentious debate in moral and political philosophy. You just say: this topic is underexplored.

Below, we'll argue that a lot of what people aim to capture with the ideal/ non-ideal distinction is unimportant or incoherent, and we suggest that the remaining sensible parts reduce to various instances of **Underexplored**.

### **Use 1: Full-Compliance vs Partial-Compliance (Rawls)**

Much of the terminological confusions surrounding 'ideal/non-ideal theory' can be traced back to the political philosopher John Rawls. In his book *A Theory of Justice*, Rawls uses these terms to mark two very different distinctions. The first distinction is between theories that assume full compliance with the demands of justice and those that don't make that assumption.<sup>3</sup>

Our question is whether this distinction is usefully applied to work in philosophy of language. First note the most obvious discrepancy: semantic, syntactic and pragmatic theories are not theories of justice and so there's no distinction there between those semantic, syntactic, or pragmatic theories that assume full compliance with the demands of justice and those that don't. Whether people are in full or partial compliance with the demands of justice in Rawls's sense is just irrelevant.

To make the distinction between ideal and non-ideal theory useful when thinking about philosophy of language, we would need move away from how Rawls uses those terms. We have to find a substitute for 'compliance with the demands of justice' that's relevant to the study of language. It's entirely non-trivial to come up with that. Here a possibility: A theory of meaning will make assumptions about how speakers behave and what semantic and pragmatic rules they follow. Maybe the analogue of Rawls's 'full compliance' is the assumption that *everyone* always follows all these rules fully. But the details of this is extremely hard to work out: 'salmon' denotes salmons in English. Is 'full compliance' then to be understood as the claim that all English speakers always use 'salmon' to denote salmon? Are all English speakers in full

---

<sup>3</sup> Stemplowska and Swift points out that Rawls gives the impression that the distinction between ideal and non-ideal theory simply reduces to a whether a theory assumes strict or only partial compliance: "Thus, he speaks of "[i]deal, or strict compliance, theory..." (2001: 13) and, even more strikingly, introduces the concept of ideal theory for the first time exclusively with reference to the problem of compliance (1999a: 7-8)."

compliance with the rigidity of proper names? Insofar as these questions make sense at all, the answers are not useful for dividing up work done in philosophy of language. We have debates of about whether there are natural languages or just micro languages, how much of language is context sensitive, how context sensitivity works, etc. Adding the 'ideal/ non-ideal' distinction will do nothing to illuminate or move those debates forward<sup>5</sup>.

Maybe the more plausible illustration would be variations on Grice's cooperative principles. Grice says:

We might then formulate a rough general principle which participants will be expected (*ceteris paribus*) to observe, namely: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. One might label this the Cooperative Principle. (1975: 45)

Now we could create an analogue of Rawls' principle as follows:

Griceans assume full compliance with the demands of the cooperative principle. Hence they're doing ideal philosophy of language.

The problem with this way of extending the terminology is that there's no one in the Gricean tradition who ever assumed *full* compliance with this or related principles. One of the core points of Gricean pragmatics is that people both flout and violate these maxims all the time. No one in the Gricean tradition thought that there were no liars or that people don't use language for deceptive and evil ends. Griceans try to describe what happens when speakers follow the cooperative principle and they think some people sometimes do that. The Gricean also predict that if we're in a situation in which it's known that speakers are being *uncooperative*, a lot of the usual implicatures will go away. And in fact this seems right. (Scalar implicatures, for example, can go away in confrontational courtroom situations.) So the Gricean not only doesn't assume that the normative principle is always being followed—they actually have a theory that says the right thing about cases in which the normative principle isn't being followed.

More generally, if we take 'full compliance' to mean that we're theorizing only about people with certain (desirable) features—people who are acting well, following the rules, being cooperative, knowing the meanings of words, etc., then (a) it's not clear what the "desirable" features are in the linguistic case<sup>6</sup> and (b) the theories *aren't* assuming that everyone has the desirable features. Minimally they're saying what happens when language is used by people with the desirable features. More expansively, they might predict that when language is used by people without those features, some of those things don't happen (as with the implicatures).

---

<sup>5</sup> We also have the performance/competence distinction; thus we have compliance at the level of performance (something we never achieve, even for idiolects) and at the level of competence.

<sup>6</sup> Is it "desirable" to apply a word only to things described by its satisfaction conditions, so that metaphorical speakers are acting undesirably?

Or they might say that when language is used by people without those features, there are no systematic consequences to characterize.

Now you might think that's all well and good, but also think that not *enough* effort has been spent exploring what happens when people aren't in full compliance. While it might be true that Grice has something to say about non-cooperative behavior, maybe he and the tradition after him didn't focus enough on, say, lying, misleading, bullshitting or various other manipulative uses of language. Below we will show that this complaint is unfounded, but let's grant it for the sake of argument. Note that what we then have is an instance of **Underexplored**. We have already established that in order to make **Underexplored** complaints, we don't need the ideal/non-ideal terminology.<sup>7</sup>

## Use 2: Ideal Theory Assumes Favourable Circumstances (Rawls)

The second way in which Rawls uses the ideal/non-ideal distinction is to contrast theories of justice that assume that the natural and historical conditions are favourable with those that don't. According to Rawls, the conditions are favourable if the economic and social conditions are sufficiently developed to realize justice.

Again, the political version of this doesn't apply in any immediate way to work in philosophy of language. Consider the claim that proper names are rigid designators. That claim makes no assumptions about justice at all. We would need a broader sense of 'favourable'. Maybe the relevant distinction would be between theories that assume 'favourable linguistic conditions' and those that don't (where the former are Ideal Philosophy of Language and the latter Non-Ideal). Does Kripke assume that linguistic conditions are favourable for achieving rigidity? For much the same reasons we gave in the previous section, it's hard to know what that question means. If proper names are rigid, then, yes, in some sense the linguistic conditions are favourable towards rigidity. Again, that's just to say that names are rigid and so the world is such that it makes proper names rigid. To contrast theories that assume 'favourable' conditions with those that don't is just to contrast theories that accept rigidity and those that don't.

As in the case of the 'full compliance' reading of 'ideal', one might suspect that the Gricean principles of cooperation are better candidates for spelling out some linguistic analogue of 'favourable' conditions. Does the Gricean tradition (with its focus on e.g., the Cooperative Principle) assume favourable linguistic conditions? Well, Grice thinks that actual people often cooperate in conversations and he has a theory of what happens when they do and what happens when they don't. That's just to say that he thinks he gives a correct description of the activity some speakers engage in some of the time. Of course, for activity A to take place, the world has to be such that A can take place in it. It is hard to see that an appeal to 'favourable

---

<sup>7</sup> Note that the explored-underexplored distinction doesn't even map accidentally onto the ideal/non-ideal distinction: free choice permission in ability modals is underexplored—that's got no connection with how 'ideal' anyone or anything is.

conditions' will do much to illuminate that work or help to contrast it with an alternative, non-ideal philosophy of language. The non-favourable conditions would be when people violate or flout the maxims, but that is something Grice talks extensively about. Again, maybe there some kinds of non-cooperation (e.g., Frankfurtian bullshitting) that are underexplored, but that's just an instance of **Underexplored**—not something we need a new distinction to characterize.

Here is another illustration of why assumptions about 'favourable conditions' aren't useful for characterizing work in philosophy of language. Imagine someone noticing that Grice and others who theorize about communication never talk about how much noise there is in the real world. In the examples Griceans discuss, speakers are never next to jackhammers or other things that disrupt communication. Griceans tend to just ignore noise levels. Now ask: is that a sense in which the Gricean tradition in philosophy of language has been assuming 'favourable' conditions (without even mentioning it)? Are they involved in theorizing that idealizes away from the real world (filled as it is with noise)? Is that a sense in which we can say that philosophy of language is ideal theory and that we could introduce a non-ideal theory that's more sensitive to auditory interference (and illegible handwriting etc.)? Surely the answer here is no. That would just be silly. Names are still rigid even when spoken next to jackhammers. Speakers try to generate implicatures in noisy rooms, it's just that they sometimes fail because they can't hear each other. Then there is another theory—the theory of auditory perception—that will explain when people can hear each other and when they cannot. That turns out to be a good division of labor because those with expertise in syntax, semantics and pragmatics don't know very much about ears and jackhammers. In sum: We don't need the distinction between ideal and non-ideal theory to explain why we abstract from noise levels. This second Rawlsian way of distinguishing between ideal and non-ideal theory doesn't apply in any interesting way to work in philosophy of language.

### **Use 3: Utopian vs. (More or Less) Realistic Theory (Valentini)**

Laura Valentini (2012) articulates a third use of 'ideal vs non-ideal theory' (albeit it one very close to Use 2 above). She says that it is sometimes is used to distinguish:

... the types of feasibility constraints we might take into account when designing normative principles. ... First, we need to distinguish between 'fully utopian' theories, which altogether reject the need to place feasibility constraints on principles of justice, and 'realistic' theories, which accept some such constraints. Second, among the latter group of theories, we need to distinguish between more or less realistic ones, depending on what kinds of real-world constraints they factor into the design of normative principles. (2012: 656-7)

Valentini uses G. A. Cohen to illustrate the Utopian view according to which justice is 'timeless (fact-free) value, perhaps akin to a Platonic ideal.' (2012: 657) Whether we can achieve it or not is irrelevant and figuring out what justice is, won't tell us what to do to get there, according to



Cohen. What would be the linguistic analogue of this kind of ideal theory? What would be a linguistic utopia? It would have to be a theory that doesn't place 'feasibility constraints' on a theory of meaning and communication. But no such theory has ever been proposed, so it's a pointless category. No philosopher of language has tried to describe an ideal language unconstrained by *any* facts about how people can use such a language. The various theories that spring out the kind of process described in section 1 are all constrained by facts about real languages and their speakers. Even theorists who constructed improved languages (e.g., Frege and Carnap) did it in order to help people—real people—think and reason better.

Turning to the distinction between the *more or less* realistic theories, it is also hard to see how that adds anything to what we already have in section 1. Theories of the meaning and effects of utterances are constrained by whatever factors the theorist choose to include. She will always abstract from some features of the real world (that is in the nature of theorizing) and focus on some at the expense of others (again, that's in the nature of theorizing). As above, we don't need the 'ideal/non-ideal theory' distinction to point that out.

#### **Use 4: Idealized Models: “an ideal—in the sense of an exemplar—model of how P should work” (Mills)**

The final attempt to distinguish ideal from non-ideal theory comes from an influential paper by Charles Mills (2005). He distinguishes between two ways of constructing models to describe the behaviour of a given phenomenon P:

*Ideal-descriptive models.* Such models purport to be descriptive of P's crucial aspects (its essential nature) and how it actually works (its basic dynamics). An ideal-descriptive model has to abstract away from certain features of P: one will make simplifying assumptions, based on what one takes the most important features of P to be, and include certain features while omitting others.

*Ideal-idealized models:* such models provide an exemplar of what an ideal P *should* be like. The “should” here will in general not necessarily be a moral “should,” but may involve norms of a technical functionalist kind (an ideal vacuum cleaner, an ideal factory farm, an ideal digestive system, and so on) or just limiting assumptions convenient for the purposes of mathematization and calculation (an ideal gas, a perfect vacuum, a frictionless plane, a resistance-free conductor).

Is this distinction useful for categorizing work in philosophy of language? First, we should give Mills credit for trying to articulate a more general distinction—it's not meant to be applicable just to work in moral and political philosophy. As a result, this is potentially more directly applicable to, for example, philosophy of language. That said, we also think that the attempted generalization is problematic because the distinction between the two kinds of theorizing is unclear. Here is what we have in mind: Ideal-descriptive models abstract away from various features of the world. Which features? In order to not be a sub-set of ideal-ideal-models, these

features must not be delineated in the broad sense of 'ideal' or 'normative' that is characteristic of the ideal-ideal models. It is hard to see how to avoid that (i.e., it is hard to avoid appealing to something normative when picking out the features that ideal-descriptive theories abstract away from). If we start classifying properties of, say, vacuum cleaners, as important and not so important (as we are doing in an ideal-descriptive model), the natural way to do that is to think about what's required for a *well-functioning* vacuum cleaner. The natural thought is to proceed as follows:

We abstract away from lots of the *deficiencies* that plague most vacuum cleaners. The ideal-descriptive model of a vacuum cleaner will exclude those that have got gum stuck inside because they are *defective*. For the same reason, they will exclude those that are cracked or have any number of potential *defects* that are found in real world vacuum cleaners. Abstraction is typically towards non-defective, *non-functioning* exemplars.

That, however, but that can't be right because we are then introducing the normative component that's supposed to be distinctive of the ideal-ideal models (and distinguish those from the ideal-descriptive models). The challenge, therefore, is to make systematic sense of the form of abstraction involved in ideal-descriptive models, without appeal to the broad sense of 'should' the ideal-descriptive models involve. We are not claiming this *cannot* be done, but if it is done, we suspect there are very few real cases left of pure ideal-descriptive models: it's an abstraction with few real instantiations. That's our first concern with Mills picture, but it won't play a central role in what follows.

Mills goes on to argue that in moral and political philosophy, ideal-idealized models are deeply problematic. In what follows, we'll spend a bit of time outlining the arguments that Mills gives for this. We do that because we suspect some readers are inclined to reason as follows: *Mills is right about the bad effects of ideal theory in moral and political philosophy and 'idealized philosophy of language' is problematic for some of the same reasons*. Our conclusion will be as before: it's pointless to classify philosophy of language as ideal or non-ideal in this sense, and there's nothing problematic here.

The problem, according to Mills, with ideal-idealized models is, "...the exploration of the ideal as an end in itself without ever turning to the question of what is morally required in the context of the radically deviant *non-ideal* actuality" (Pateman and Mills 2007: 118). Mills says:

If we start from what is presumably the uncontroversial premise that the ultimate point of ethics is to guide our actions and make ourselves better people and the world a better place, then the framework above will not only be unhelpful, but will in certain respects be deeply antithetical to the proper goal of theoretical ethics as an enterprise. In modeling humans, human capacities, human interaction, human institutions, and human society on ideal-as-idealized-models, in never exploring how deeply different this is from ideal-as-descriptive-models, we are abstracting away from realities crucial to our comprehension of the actual workings of injustice in human interactions and social institutions, and thereby guaranteeing that the ideal-as-idealized-model will never be achieved. (2005:170)

Okay, back to philosophy of language. Let's try to mimic Mills's rhetorical flourishes applying them to theories of meaning and communication. You might try the following:

If we start from what is presumably the uncontroversial premise that the ultimate point of philosophy of language and communication is to guide our speech and make ourselves better speakers, then the ideal-idealized philosophy of language will not only be unhelpful, but will in certain respects be deeply antithetical to the proper goal of philosophy of language as an enterprise. In modeling human speech on ideal-as-idealized-models and in never exploring how deeply different this is from ideal-as-descriptive-models, we are abstracting away from realities crucial to our comprehension of the actual workings of speech and communication and thereby guaranteeing that the ideal-as-idealized-model will never be achieved.

This type of argument is deeply flawed. First, the goal of philosophy of language is not to make us better speakers or communicators.<sup>8</sup> That much is, we take it, common ground here. Second, as we have seen, philosophers of language have not presented ideal-idealized models of how language ought to be. They've tried to describe how languages actually are and how speakers actually behave. In so doing they by necessity abstract from certain feature of real speakers, but that's essential to all theorizing (as Mills recognizes). Russell doesn't say that descriptions *ought* to be quantifiers. He says that they are. Kripke doesn't claim that names ought to be rigid designators. He claims that they are. Grice doesn't claim that people ought to follow the cooperative principle. He claims that some people sometimes do (and that if you want to exchange information, the maxims of conversation are useful). So (putting aside the problems with understanding Mills' distinction between ideal-descriptive and ideal-idealized models) the objection to ideal-idealized models doesn't work when applied to philosophy of language.<sup>9</sup>

Mills goes further and argues that ideal-idealized models in political philosophy are ideologies that reflect the interests of middle-to-upper-class white males (i.e., people such as John Rawls). He says:

Ideal theory, I would contend, is really an ideology, a distortional complex of ideas, values, norms, and beliefs that reflects the nonrepresentative interests and experiences of a small minority of the national population—middle-to-upper-class white males—who are hugely over-represented in the professional philosophical population. (2005: 172)

---

<sup>8</sup> Just as the goal of ethics is not to guide our actions or to make us better people. It's to understand the nature of morality and justice.

<sup>9</sup> We should add that even if there were ideal-idealized models in philosophy of language, they would have no impact on whether we achieve the ideal because theories in philosophy of language and linguistics have absolutely no impact how speech develops (just as political philosophers have no impact whatsoever on whether societies become more just (or, more cautiously, no more impact than a butterfly flapping its little wings)).

You could try to mimic this as applied to theories of meaning and communication.

*Question:* Are theories of meaning and communication a distortional complex of ideas, values, norms, and beliefs that reflects the nonrepresentative interests and experiences of middle-to-upper-class white males who are hugely over-represented among philosophers of language and linguists?

Again, we take it the answer here is fairly non-controversial:

*Answer:* No. People who are not middle-to-upper-class white males also use names as rigid designators, definite descriptions as quantifiers, and when they talk to each other they try to be cooperative.

Before we leave Mills behind, we'll pause for a moment and consider a kind of objection that might easily be misunderstood as being aligned with Mills's.

*Possible objection:* You keep saying that it's not just middle-to-upper-class white males who use names as rigid designators—all English speakers do. But experimental philosophers have given us powerful evidence that non-western cultures don't use names as rigid designators (Machery et al. 2004). Isn't this exactly the kind of thing that Mills had in mind? It shows that the nonrepresentative interests and experiences of western academics has distorted core elements of philosophy of language. Doesn't this show that mainstream philosophy of language is idealized in exactly Mills's sense?

Our reply will summarize core points from above: the kind of concern outlined in the objection is exactly the kind of data that our central principle **Underexplored** is perfectly placed to account for. The complaint is that there's important data that some theorists have overlooked and that the reason they have overlooked it is because the majority of theorists have certain socio-cultural biases. That happens all the time. Often people will be reluctant to take those kinds of objections seriously because it challenges orthodoxy. That's just an obvious corollary of the fact that researchers are human beings with all the accompanying defects. We really don't need to import terminology from moral/political philosophy to describe that.

### **Talking Stock: Philosophy of Language and the Distinction between Ideal vs Non-Ideal Theory**

We have looked at four ways the ideal/non-ideal theory distinction has been drawn by influential moral and political philosophers. We have tried to see whether the distinction can be extended to illuminate an interesting divide in philosophy of language (between ideal theories and non-ideal theories). Our conclusion, in each case, has been that there is no useful extension. We suspend judgement about whether it's a useful distinction in moral and political philosophy (though we have to admit we have our doubts), but if it is, that has do with peculiarities of those disciplines.

With this in mind, let's consider some authors who explicitly invoke the ideal/non-ideal distinction when talking about philosophy of language. The authors of this paper provide a good illustration. In the first chapter of *Bad Language*, for example, Cappelen and Dever identify what they call "some idealizations that have guided many of the efforts to theorize about natural language, communication, and speakers" (Cappelen and Dever 2019: 1), and contrast these idealizations with "tools needed to deal with a decidedly non-ideal world" (1). Their list of idealizations includes items such as:

- Conversation and communication are fundamentally cooperative exercises.
- Speakers say only what they know.
- A common language has words with stable meanings known to all conversational participants.

But "lists of idealizations" of this sort are confused in multiple ways.

- First, as the discussion of the previous section has shown, there is no helpful sense in which these are *idealizations*. Speakers that cooperate or who say what they know, or languages with stable word meanings, aren't in any interesting sense *better* or *more properly functioning* than speakers or languages that don't. These are, if anything, *simplifying assumptions*. But simplifying assumptions are all over the place, and don't follow any systematic pattern. Sometimes we simplify by assuming that a body of information is common knowledge for the conversational participants—known by all the speakers, and known to be known by all the speakers, and so on. At other times we simplify by assuming that adjectives are all intersective. And at other times we simplify radically by considering a language that contains only a single modal of interest and some uninterpreted sentence letters.
- Second, it's not clear that work in philosophy of language really is making those particular simplifying assumptions. It's certainly true that many formal tools are developed to deal with simplified cases—an intersective semantics for adjectives really will have trouble dealing with "tall", and a material conditional semantics for "if" really will have trouble dealing with almost all natural language conditionals. But it's much less clear which pieces of theoretical machinery are going to break or misfire when speakers aren't being cooperative or when contextual information isn't common knowledge. In fact, as we've noted above, some of the standard tools look like they make exactly the right predictions in these cases (no Gricean conversational implicatures result from violations of the maxim of quantity when speakers aren't cooperative; the update effects of epistemic modals become non-trivial when contextual information has a non-S5 structure<sup>10</sup>).

---

<sup>10</sup> "To say that it is unclear what the contextual information is, is to say that different epistemically possible worlds, different sets of worlds represent the contextual information. Thus not all worlds in the context set agree on what worlds are in the context set, and the accessibility relation is not S5. Lewis's example "You can put the public interest first for once!" (Lewis 1979) is the classic example of an existential (here, practical)

- And third, even where the simplifying assumptions *are* being made, they are typically being made as steppingstones toward a more sophisticated theory capable of handling less simplified data.

### 3. Is the category ‘social and political philosophy of language’ more helpful?

Our job for this volume was to clarify the distinction between Ideal and Non-Ideal Philosophy of Language. We’ve ended up rejecting the distinction. However, maybe there’s another, closely related distinction that is useful. This volume has in its title the phrase ‘social and political philosophy of language’. What does that category pick out and what is it contrasted with?

At the risk of appearing to be incurable curmudgeons, we are going to also reject that category as fairly useless. What could it possibly pick out? Maybe it’s an attempt to pick out speech by people who are politicians or people who talk about political/social topics. If that’s the remit of social-political philosophy of language, then it should be concerned with sentences like:

- “We should increase the sale tax on cigarettes because that would give the city more money for schools and it could reduce cancer rates.”
- “The price of parking on city streets should be increased because it is not fair that public property should be rented to car owners for very little money.”

That’s what the vast majority of ‘political speech’ is like. It’s just ‘ordinary speech’. It’s not special in any way. It doesn’t constitute a distinct subset of speech. The above are just normal sentences uttered by normal speakers who have certain interests or have certain jobs. The same point applies if the claim is that ‘social and political philosophy of language’ picks out words or constructions or speech acts that are politically relevant. The act of saying: “The meeting starts on Thursday at 5pm” is politically relevant, if the meeting is about politics, but there’s nothing semantically distinctive about the sentence or the act.

### 4. A Final Proposal: Two Lists

It turns out to be impossible to find definitions of ‘ideal/non-ideal’ or ‘social and political’ that will do useful work in carving out interesting subsets of work in philosophy of language. One way to bypass that problem is to just have a list of what people think of as falling into these

---

modal having a non-trivial update effect because of lack of common knowledge of the contextual options. (Willer 2013) gives a general model of update effects of epistemic modals under conditions of contextual uncertainty and shows how that model captures many standard conversational uses of those modals; the tools of Dynamic Epistemic Logic (van Benthem 2011) allow modeling of epistemic updates in a wide range of non-S5 situations involving epistemic interactions of agents under conditions of limited shared knowledge.”

categories—without asking them to justify or define them. Here’s a conjecture: if you were to ask a group of contemporary philosopher of language to produce two lists, one with topics from ‘social and political or non-ideal philosophy of language’ and one from ‘ideal or not-social-or-political philosophy of language’, they would probably know what to do and they would end up with fairly similar lists:

#### List 1

- Slurs,
- Bullshitting,
- Lying,
- Insincerity
- Silencing
- Manipulation and propaganda
- The language of race, gender, and disability
- Lexical effects
- Code words
- Conceptual engineering
- Generics

#### List 2

- Rigidity of proper names and natural kind terms
- Gricean maxims and the production of conversational implicatures
- De re and de dicto readings of attitude reports
- Externalism and naturalness in metasemantics
- Generalized quantifiers
- Presupposition projection
- Update conditions of embedded questions
- Donkey anaphora
- Indeterminacy of translation and underdetermination of meaning
- Semantics of vague expressions
- Lexical semantics

Suppose our conjecture is true: there would, for example, be somewhat broad consensus that the topics on List 1 belong in this volume and that the topics on List 2 don’t.<sup>11</sup> Suppose also that we are right that List 1 topics have no unity. In particular: they don’t share characteristics that List 2 topics lack. What are we to conclude from this? At least this: List 1 is a more or less random collection of topics with no internal unity that for various hard to understand sociological reasons tend to be lumped together. This might seem surprising (and disappointing for those who have bought this volume expecting the topics covered to have a certain kind of unity). However, it’s not all that unusual. We freely operate with distinctions like ‘analytic vs

---

<sup>11</sup> The conjecture is true: the topics on List 1 are in this volume and it has gone through a thorough refereeing process.

continental philosophy' where there's also no unity, but whereby some random convention philosophers would know how to make lists. Maybe even 'philosophy' itself is like that: a list of topics with no thematic or methodological unity. More generally, our classifications and divisions are often random and lack internal coherence in this way and it's important to emphasize that this doesn't make topics on the lists less interesting or less worthy of being explored.

However, divisions created in these random ways can sometimes have the effect of blinding us to important interconnections. In the case of List 1 and List 2, that would be very unfortunate because the topics on the two lists are so intimately intertwined. We end with some brief case-studies that illustrate this.

#### Case-Study 1: From Kripke to Scorekeeping in a Pornographic Language Game

Consider one historical line of development in philosophy of language. In the 1960s, work by Saul Kripke on modal logic and by Arthur Prior on tense logic led to the development of a formal toolkit of methods involving truth at an index of evaluation. In the late 1960s, Hans Kamp, working with that toolkit, uses data about the temporal and modal indexicals "now" and "actually" to motivate extending these methods to a two-dimensional notion of truth relative to a pair of indices. David Kaplan then in the early 1970s provides a philosophical setting for the formal methods by introducing the character/content distinction and thus allowing for careful semantic analysis of context-sensitivity and careful representation of contexts as agent-time-position-world tuples. Then in the late 1970s, David Lewis in "Scorekeeping in a Language Game" generalizes the Kaplanian notion of a context to include such features as commonly shared bodies of information, conversationally-licensed permissions, and plans. Lewis's "scorekeeping" picture then gets deployed, in the 1990s and early 2000s, by Langton and West in "Scorekeeping in a Pornographic Language Game" and by Mary Kate McGowan in her work on oppressive speech to give detailed theoretical stories about particular ways in which the structures of language interact with power dynamics in society.

#### Case-Study 2: Slurs and Race/Gender Terms

In the 1940s and 1950s, work by W.V.O. Quine and Ruth Barcan Marcus on the interaction between modal operators and quantifiers led especially Marcus to an interest in the way that modality shapes the best theory of reference. Further considerations along these lines led Kripke in *Naming and Necessity* in 1970 to defend a broadly direct reference view of names via a metasemantic mechanism that placed reference fixing in the hands of external causal mechanisms. This initial externalist idea in semantics is further developed by Hilary Putnam in the early 1970s, and is then given a specifically social development by Burge's work in the late 1970s and early 1980s emphasizing the way that the semantic burden of meaning determination is distributed across a speech community, rather than being individualized. The resulting externalist picture, with its easy accommodation of a speaker alienated from their



own conceptual repertoire, is then used by Sally Haslanger in “What Are We Talking About? The Semantics and Politics of Social Kinds” to give an account of the meaning of gender and race terms that fits with her ameliorative project and by Chris Hom in “The Semantics of Racial Epithets” to produce a theory of the normative content of slurs.

### Case-Study 3: The Essentializing Effects of Generics

Bertrand Russell’s investigation of type theory as a solution to the semantic paradoxes leads to the development of categorial grammar by Kazimierz Ajdukiewicz and the lambda calculus by Alonzo Church. These formal tools are then taken up by Richard Montague in the late 1960s and very early 1970s. In “The Proper Treatment of Quantification in English”, he uses a rich type hierarchy in giving a semantic theory for a fragment of English. The use of higher types in Montague semantics then leads to Greg Carlson investigating the semantics of reference to kinds in the late 1970s, sparking work on the semantics of generics. Simultaneously, early work on the semantics of “donkey pronouns” by Peter Geach and Gareth Evans leads to David Lewis’s treatment of adverbs of quantification, which is then put to work in giving quantificational analyses of generics. The link between adverbs of quantification and modal adverbs then leads to “normal worlds” modal analyses of generics such as Nicholas Asher and Michael Morreau’s work in the 1990s. The combination of the modal element and the reference to kinds in generics then leads to work by Sally Haslanger and Sarah-Jane Leslie exploring the tendency of generic language to give rise to problematic essentializing inferences with bad social consequences.

We conclude from these case studies that the divide between List 1 and List 2 doesn’t represent a substantial or theoretically interesting divide in the field. There is of course a general movement from theories built to handle simple cases to theories that extend those earlier theories to handle more complicated cases. The items on List 1 are perhaps in general more complicated than the items on List 2 (although the comparative complexity of, for example, slurs and donkey pronouns doesn’t seem like a matter that can be easily evaluated), so there is perhaps a general tendency for theoretical work to progress from tools crafted in addressing List 2 topics to the use and extension of those tools in addressing List 1 topics. But all of this theoretical work is being done together, with tools, distinctions, and data being freely passed from one topic to another. In philosophy of language, not only does tomorrow not never come—it is already here.

Of course, this is not to say that there are not unjustly neglected topics in philosophy of language. Surely there are. Philosophy of language is, it’s worth remembering, a young field. A bit more than a century old at the most; little more than 50 years old by slightly stricter standards. It’s far too early for us even to have the slightest idea what all the important topics are, much less for us to have adequately addressed all the topics. But “unjustly neglected” doesn’t track anything like a putative ideal/non-ideal philosophy of language distinction, and indeed doesn’t look like it tracks *any* interesting distinction (other than perhaps the “hard but maybe tractable/just impossibly hard” distinction).

Back to work, then, everyone — there's much to be done.

### **Bibliography**

Cappelen, Herman & Dever, Josh (2019). *Bad Language*. OUP.

Grice, Herbert Paul (1967). *Logic and conversation*. In Paul Grice (ed.), *Studies in the Way of Words*. Harvard University Press. pp. 41-58.

Lewis, David (1979). *Scorekeeping in a Language Game*. *Journal of Philosophical Logic* 8:1, 339-359

Machery, Edouard ; Mallon, Ron ; Nichols, Shaun & Stich, Stephen (2004). *Semantics, cross-cultural style*. *Cognition* 92 (3):1-12.

Mills, Charles W. (2005). "Ideal theory" as ideology. *Hypatia* 20 (3):165-184.

Pateman, Carole & Mills, Charles (2007). *The Contract and Domination*. Polity

Rawls, John (1971). *A Theory of Justice*. Harvard University Press.

Stemplowska, Z., & Swift, A. (2012). *Ideal and nonideal theory*. *The Oxford handbook of political philosophy*, 373-389.

Valentini, Laura (2012). *Ideal vs. Non-ideal Theory: A Conceptual Map*. *Philosophy Compass* 7 (9):654-664.

van Benthem, Johan (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press.

Willer, Malte (2013). *Dynamics of Epistemic Modality*. *Philosophical Review* 122:1, 45-92

Zwarts, J. (2017). *Spatial semantics: Modeling the meaning of prepositions*. *Language and Linguistics Compass*, 11(5), e12241.