Groundwork for a Pragmatics for Formal Languages

David Kashtan

February 22, 2021

Abstract: *The use-mention distinction is elaborated into a four-way distinction between use, formal mention, material mention and pragmatic mention. The notion of pragmatic mention is motivated through the problem of monsters in Kaplanian indexical semantics. It is then formalized and applied in an account of schemata in formalized languages.*

Keywords: Use vs. mention, Semiotics, Semantics, Pragmatics, Formal languages, Truth, Direct reference, Schemata, Indexicals

1   Introduction

1.1   The basic intuition

It is not often recognized that the use-mention distinction is not a two-way, not even a three-way, but really a four-way distinction.[1] To begin, look at the following sentence:

(1)   Cats have four legs and 'cats' has four letters.

In its first occurrence in (1), the string 'cats' is *used*; in the second it is *mentioned*.

(2)   'cats' (in English) denotes cats.

In its mentioned occurrence in (1), only the phonological properties of 'cats' are appealed to – it is mentioned as a mere string of phonemes.[2] In (2), by contrast, it is mentioned as an *interpreted* string of phonemes, as a genuine word in a language. Accordingly, we distinguish within the category of mention

---

2. I use the term 'phoneme' and its derivatives liberally to stand for the perceptible and producible aspect of a linguistic object generally, so as to include, in particular, letters of the alphabet.

between, on the one hand, phonological or *formal* mention and, on the other hand, semantic or *material* mention.[3]

But now observe:

(3)   In its first occurrence in (1), the string 'cats' is used.

In this last sentence, the string 'cats' is once again mentioned, but this time it is mentioned *as being used*. This, I submit, is a third mode of mention, distinct both from the formal and from the material mode. Since it targets the *use* of a string, I will label it *pragmatic mention*. With it, we are up to four "modes" in which an expression can occur in discourse: used or mentioned, and when mentioned, either formally, materially, or pragmatically.

My distinction between formal and material mention shouldn't knock anyone out of their chair, and yet it is not always recognized with sufficient clarity. For example Kripke, in his famous (1975) treatment of truth and the semantic paradoxes, remarks that "Gödel put the issue of the legitimacy of self-referential sentences beyond doubt" (p.692). This remark is correct if formal we understand 'self-reference' in terms of formal mention, but quite off the mark if material mention is meant. For if, as seems reasonable to suppose, to materially mention an expression implies being able to state its truth conditions, then the problems that beset the concept of truth, in particular the so-called semantic paradoxes, carry over to material mention. To solve the issue of *material* self-referential sentences is then tantamount to solving the liar paradox, which Gödel of course has not done (and which Kripke in that very paper is in fact attempting to do).

By contrast, the notion of pragmatic mention is, I think, new. The purpose of the present paper is to advertise it, develop it somewhat, and demonstrate its importance. The strength of pragmatic mention, it will be argued, lies in its weakness. Appealing to an occasion of *use* allows us to tap into an expression's meaning without the theoretical (and especially ontological) commitments that full semantic interpretation usually incurs. Technically, pragmatic mention will be developed in the form of a semantic framework based on pragmatic "meaning-giving" directives, which don't amount to full-blown meaning clauses, but which manage to do of the latter's work. The philosophical underpinning for directive semantics will be given in §2, in the context of Kaplan's theory of indexical reference, and in particular his "prohibition on monsters". In §3 I generalize and adapt directive semantics to formalized languages, and in §4 I argue that the adaptation is worthwhile by posing a problem in formal language theory, the problem of schemata, and solving it using directive semantics. In the remainder of the present section I develop the intuitive notion a little further and lay down some methodology.

1.2   Ontology and methodology

Language *use* is a kind of action, and it involves a specific kind of agency, namely linguistic agency or competence. Using an expression e implies competence with

---

3. There are intermediate cases, such as when a string is mentioned as syntactically or logically structured, though without interpretation of its content. I don't discuss these cases.

the language of e.[4] By contrast, mentioning e does not imply competence with the language of e anymore than mentioning a piano, say, implies competence on the piano. There are expressions we can mention though we can be sure never to be able to use, such as infinitely long formulas, or expressions that are ontologically identified with mathematical objects.[5]

When we *mention* some expression e, there will be another, typically distinct, expression f, often in a language different from e's, which we will be *using* in order to mention e. Therefore mentioning e *does* involve linguistic competence, but the language in question is that of f, not of e. In all cases, I will refer to the language, competence with which is required, as the *matrix language*. Note that there will usually not be a unique such language. If I can mention e from within a language L, then, for example, every extension of L will mention e. More closely, then, the term *matrix language* denotes something like the weakest language, competence with which is required for the case at hand.

The four modes of the use-mention distinction make different claims on their respective matrix languages. The concept of matrix language can therefore serve as a point of reference with which to compare the commitments inherent in each mode. In general, this idea is captured in the following methodological maxim:

Postulate (Matrix Language) *Identify the language of the expressions that you're using.*

This principle is applied, for example, by Kripke in his aforementioned (1975), when he notes (p. 714) that the language from within which he develops his truth theory (his matrix language) cannot be identical with the language of the truth predicate he defines, and therefore that "the ghost of the Tarski hierarchy is still with us".

This principle only makes sense if we have a definite way to compare languages. We will compare them in terms of their *expressive resources*. The languages I will be concerned with primarily are fully interpreted formalized first-order languages of the usual (extensional) kind (henceforth *standard languages*). The only notion in this paper that will not be expressible in a standard language, and that will require the definition of a new kind of language, is that of pragmatic mention. Following Quine, I distinguish in standard languages two categories of expressive resource: (a) *ontology*, or the range of objects that can be referred to from within the language; and (b) *ideology*, or the range of concepts that the language can express. In standard languages, ontology is identified with the domain of quantification and ideology amounts to the class of definable subsets of that domain.

---

4. Simply *voicing* a string of phonemes parrot-like doesn't amount to *using* it, even if that string happens to be meaningful in some language or other.

5. Gödel used numbers to *code* expressions, but many contemporary authors define expressions to *be* numbers or sets to begin with, or at least they abstract from the ontological question altogether. See, e.g., Enderton (1972, 15), for whom linguistic expressions can be sequences of marbles. This approach is fine for the purposes of mathematical logic, but will not do for our purposes, since it doesn't account for the possibility of language *use*. For while it is reasonably clear what it means to put a phonological string to (linguistic) use, the same cannot be said for numbers or marbles. (Compare Carnap (1942, 3) on flags and rockets.)

The action in this paper is confined to ontology. Each of the four modes of the use-mention distinction makes a different ontological claim on its matrix language. Consider use vs. formal mention. The *use* of, say, the word 'cats' (in its English meaning) implies competence with a language L_c in which cats, the animal, can be designated. Therefore L_c should have cats in its ontology.[6] There is no presumption, however, that the *string* 'cats', that concatenation of graphemes or phonemes, should be included in L_c's ontology. Strings are just not part of the subject matter of L_c. So much for *using* 'cats'. In the case of *formal mention*, things are the other way around. Formal mention of 'cats' implies a language L_{str} with an ontology containing the strings, though in the general case devoid of cats. Use and formal mention are therefore independent of one another: I can *use* a string e in a language that can't mention it, and I can *mention e* using a language of which it is not a well-formed expression.

Mentioning 'cats' *materially* targets not just the string, but also its interpretation. Let 'cats' be a meaningful word in the language L_c, meaning more or less what it means in English. I can mention this word materially only from within a matrix language M_c that fulfills the following two conditions. First, M_c refers both to the string 'cats' and to the animal cats. But this won't be enough, since merely being able to refer both to 'cats' and to cats is perfectly compatible with 'cats' referring to dogs. The second condition is that we can express in M_c the fact that the string means what it means. A direct way to go about it is to postulate some object that will function as the string's meaning. We can then formalize material mention of an expression e as reference to the pair $\langle s_e, m_e \rangle$ consisting, respectively, of the string of e and the meaning of e. However, we wouldn't want to stipulate this without a more concrete idea as to what kind of objects meanings are, and there is no generally agreed-upon theory of such meaning objects. The most common approach, which equates meanings with functions from possible worlds to extensions, although useful, clearly fails to capture meaning in its full generality. In fact, there are prominent philosophers in the past, present and future that wish to do away with individual meanings altogether. In short, the situation with meaning objects seems to me too precarious to use them to model material mention.

Putting individual meanings aside, what is clear is that a string is meaningful when understood as part of a particular language. For example, of the string 'Gift' on its own (i.e. formally), we can say that it has four letters, one vowel, one velar consonant, etc. But in order to say that it denotes something that one gives on a birthday (as it does in English), rather than a substance that can cause bodily harm (as it does in German), one has to mention it *as belonging to English*.[7] This can be done either by saying so explicitly, or, more often, by relying on the circumstances in which the expression is mentioned, for example

---

6. I'm being loose here in the interest of smooth exposition. An English common noun such as 'cat' is standardly formalized as a predicate. A predicate makes no claim on the ontology of the language in question. So, strictly speaking, if standardly formalized the use of the word 'cats' doesn't imply an ontology of cats.

7. Common nouns in German are always capitalized , so there is after all a formal difference, but let's ignore that.

the language of the conversation (though this is not a foolproof method). In practice, strings that have different meanings in different languages very rarely cause any trouble, but theoretically it makes sense to say that material mention of an expression contains implicit mention of its language, or at least that reference to the language (together with formal mention of an expression) is sufficient for the material mention of that expression. Consequently, we will identify the material mention of an expression e with reference to a pair $\langle s_e, L_e \rangle$, consisting respectively of the string of e and the language of e. This makes material mention more holistic than it would be if we allowed ourselves individual meanings, but at least it doesn't depend on the contentious concept of individual meanings. Not that the concept of language is free of philosophical difficulty, but unlike the idea of individual meanings, it seems unavoidable.[8]

Having limited the scope of our discussion to standard formalized languages, we can identify a language with a model-theoretic interpretation function, one that specifies a domain of quantification D, maps primitive expressions to aspects of D in the usual ways (individual constants to members, unary predicates to subsets, etc.), and defines truth and satisfaction conditions for well-formed formulas, again in the usual ways.[9] Ontologically, the upshot is that the matrix language in the case of the material mention of an expression e has to have the very language of e in its ontology.

What ontological commitments does pragmatic mention incur? Here the situation is tricky. Unlike in the case of formal mention, and like in the case of material mention, when we mention 'cats' *as used*, we go beyond the mere string and make some kind of reference to the language *in which* it is used. However, unlike in the case of material mention, we don't have to include that language in the ontology of the matrix language. Imagine you and I overhear a conversation in which the following sentence is uttered:

(4)   koty jedzą myszy.

Neither of us understands this sentence (let us assume), but the situation makes it as certain as possible that something or other was expressed. We are in a

---

8. An anonymous referee has remarked to me that when we refer to an object under a mode of presentation, it is incorrect to say that we are referring to the mode of presentation as well. According to a plausible analysis of material mention of an expression e, what is strictly speaking referred to is just the string s_e, and the language L_e belongs rather to the mode of presentation. It follows that in material mention, the language of the mentioned expression is not strictly speaking referred to, *contra* the analysis in the text. This objection seems to me by and large correct, and in §3, the assumption that material mention involves proper reference to a language is dropped in favor of a weaker analysis. I thank the referee for pressing me to clarify this.

Note, however, that it is unclear to what extent the analysis in terms of modes of presentation applies here. Modes of presentation, and related notions such as aspectual shape (J. Searle (1992)) and cognitive fix (Wettstein (1986), Korta and Perry (2011)), belong in theories that address issues of cognitive significance, mental intentionality, communication, and the like. These are not the issues addressed in this paper, and their connection to the latter is not obvious. More on this in §1.3.

9. A more detailed definition is given in §3.

position to mention the sentence (4), not just as an uninterpreted string, but as a string expressing some thought, though we know not which. Since by assumption we are not in possession of a semantic theory for its language, we are not in a position to mention the sentence materially. Yet we are quite sure that the utterance is linguistic and meaningful, so our reference to it is different from merely formal mention. Pragmatic mention thus occupies a middle ground, as it were, between formal and material mention. The problem is to determine what this middle ground involves. Take the word 'koty'. Though we don't know *what* it denotes, we can refer to *whatever* it denotes by referring to its speaker. Thus we can allude to "whatever the string 'koty' means in the mouth of the speaker of (4)", referring blindly and mediately to something that might not even be present in our ontology. If the reference of "koty" need not be in our ontology, and if (as seems reasonable to assume), having a language in our ontology implies including its ontology in ours, then it follows that in order to mention an expression e pragmatically, we need not have its language in our ontology. We are, so to speak, relaying ontological commitment to the other speaker.

On analogy with the case of material mention, we can think of pragmatic mention of an expression e as reference to a pair $\langle s_e, L_e \rangle$ of the string and language of e, except that reference to L_e is no longer understood as reference properly so-called, but as some kind of "abstract" reference which does not carry ontological commitment. In this way the problem is reduced to that of expounding and accounting for this special mode of "abstract" reference. We may assume that the referent is still a language in the usual sense, i.e. an interpretation function, but we do not assume that this interpretation function is a member of the ontology of the matrix language. Thus pragmatic mention allows reference to a meaningful string without ontological commitment to the language it is in. As will be argued in §4, this is a device of great value, and the task is how to substantiate it philosophically. This will be done in §2.

1.3   Semiotics

Before we go there, however, a couple of words are in order about the kind of project we are engaged in. It will be useful to set it within the conceptual framework of Carnap's (1942) classic exposition of semiotics. The term "semiotics" comes from C.S. Peirce, where it denotes the general science of signification. In Carnap it is used for the study of specifically linguistic signification. Within semiotics Carnap distinguishes between *syntax*, which is the study of the form of linguistic expressions, in abstraction from their content and their use; *semantics*, which studies relations of form and content, in abstraction from use; and *pragmatics*, which studies all three dimensions. Notice the parallel with our modes of mention: Syntax mentions its objects as mere strings;[10] semantics mentions its objects as interpreted strings; and pragmatics, at least nominally, mentions expressions *as used*.

---

10. Carnap doesn't distinguish between phonology and syntax, though he should. See footnotes 3 and 5.

In addition to the three-way distinction, Carnap also draws a two-way distinction between 'pure' and 'descriptive' semiotics. Roughly, this distinction corresponds to that between formalized and natural languages, respectively. Descriptive semiotics has to do with languages as they are *given*, either empirically, or phenomenologically, or in some other way. Thus linguistics is an obvious example of descriptive semiotics, but more generally we can apply the label to the study of any linguistic or quasi-linguistic system of representation that is *given* in the appropriate sense. For example, under descriptive semiotics we can include much of the discipline of "metasemantics", which is, roughly, the metaphysics of language, insofar as it is practiced as the metaphysics of *natural* language.[11] Also studies of mental intentionality may be classed under descriptive semiotics – *semiotics* since they deal with signification as articulate as linguistic signification, and *descriptive* to the extent that they are constrained by empirical or phenomenological data.[12]

Descriptive semiotics, then, is concerned with linguistic (or quasi-linguistic, like the mental) systems of signification that are *given*. What this excludes in particular is formalized languages, since they are not given in the appropriate sense. Not that formalized languages are absent from modern descriptive semiotics. One quite central role that they play is as a model for natural language. Another is as a medium in which logically precise theories may be advanced.[13] But in these capacities they are of merely instrumental significance. Any philosophical importance they have is, on the descriptive approach, derivative, dependent on the philosophical importance of natural language.

Pure semiotics has to do, not with the description of any particular language, but with the philosophical analysis in general of semiotic notions such as syntactic connection, designation and reference, and others. It is not theoretically constrained by empirical or phenomenological facts: as Carnap puts it, it is "entirely analytic and without factual content" (op. cit., p.12). As an example of a theory of pure semantics, we can take Tarski's analysis of the concept of truth in formalized languages (see Carnap, ibid., pp.vi-vii, 27-29). The hallmark of pure semiotics is that it accords formalized languages an original, rather than derived, philosophical importance. In being the product of conscious convention rather than an extract from messy empirical reality, formalized languages are a more perspicuous and reliable medium of signification than natural languages. Consequently, the philosophical content of the semiotic notions is more clearly visible in the former than in the latter. Of course, natural language still serves as the original bed of intuitions for pure semiotics, but not as a source of empirical constraining evidence.

Putting the two-fold division of semiotics into pure and descriptive together with the three-fold division into syntax, semantics and pragmatics, we expect a six-cell matrix dividing our subject matter. Carnap himself, however, never

---

11. See, for example, the collection Burgess and Sherman (2014), esp. the introduction for a discussion of what metasemantics is.

12. Examples include J. R. Searle (1983), Korta and Perry (2011), Recanati (2016).

13. Compare Yalcin (2018).

mentions *pure pragmatics*, though he explicitly refers to the other five combinations. Moreover, he describes pragmatics as "the basis for all of linguistics" (p.13), where linguistics is for him synonymous with descriptive semiotics. This intimate connection between pragmatics and linguistics is due to the fact that linguistics proceeds by "[observing] the speaking *habits* of the people who *use* it" (ibid., emphasis added). Since pragmatics is the study of language *use*, linguistics depends on it. But it seems Carnap is here mistakenly conflating *actual* language use, as is present in empirically observed "speaking habits", and the mere *concept* of language use. Let's grant that in order for language to be available to empirical observation, it has to be actually used.[14] This doesn't preclude the possibility of studying the *notion* of language use in abstraction from any actual *instance* of use. Such a study would have be an example of *pure pragmatics*.

The subject matter of pure pragmatics will therefore be the idea of language use, in abstraction from aspects that belong to the ways in which language use is given in empirical or phenomenological reality, such as time of use, speaker, etc. The *pure* notion of use is timeless and disembodied, like the notion of use as it figures in the use-mention distinction. Therefore, just as formal mention is the mode of mention used in pure syntax, and material mention in pure semantics, the mode of mention appropriate to pure pragmatics will be what we've called pragmatic mention.

2   Indexical semantics

One philosopher who did glimpse the possibility of pure pragmatics is Bar-Hillel (1954), who identified it with (or at least superordinated it to) "the erection of indexical language-systems" (p.369). Bar-Hillel himself did not develop a theory of indexicality, but his suggestion was taken up later in Kaplan (1989). Kaplan's contribution consists largely of two parts. First, there is a semiotic thesis to the effect that indexical expressions possess a special mode of reference, namely *direct* reference. A corollary of the thesis is the famous prohibition on monsters. Second, Kaplan develops a doubly-indexed formal semantic system designed to capture the behavior of indexicals. In this section I argue that the direct reference thesis, and especially the prohibition on monsters, contains an important pure-pragmatic core, but that the double-index semantics is, from a pure-semiotic perspective, uninteresting. I replace it with a different semantics, based on the idea of meaning directives. This directive semantics will be the basis for the development, in §3, of the notion of pragmatic mention for formalized languages.[15]

---

14. Modern linguistics seems to be less dependent on this assumption, at least in theory. On Chomsky's view of language, for example, one might theoretically learn about language by studying the brain or the genome under a microscope, without observing speaking habits.

15. This section summarizes work carried out more fully in Kashtan (2018) and (2020).

2.1 Kaplan's Indexical Semantics

Before Kaplan, Bar-Hillel's suggestion of modeling indexical phenomena was taken up by Montague and others, who relativized an essentially Tarskian semantics to a parameter, called an 'index', representing a point of reference. The index represented a cluster of objects and circumstances such as a speaker, a time, etc. An indexical term would refer to one of them - 'I' would refer to the speaker, 'now' to the time, etc. If we identify the index with the context of utterance, the semantics accounts for the context-dependence of indexical reference pointed out by Bar-Hillel. However, Montague used the index parameter also in his modeling of the behavior of intensional operators. An operator such as 'always' would be defined in terms of quantification over the time dimension of the index with respect to which the expressions within the syntactic scope of the operator would be evaluated. This had the undesired effect that indexical reference would be bound by an intensional operator, if within the latter's syntactic scope, and if they both targeted the same dimension of the index. This didn't fit the phenomena. For example, temporal operators such as 'always' were never (or almost never) found to bind temporal indexicals such as 'now'. Kaplan's main formal innovation was to relativize the semantics to an additional parameter, also designating a cluster of circumstances and explicitly labeled *context of use*, so that indexicals could refer to dimensions of the context, while intensional operators would bind dimensions of the index. This made sure that indexicals are always "top scope", i.e. their interpretation is evaluated always relative to the circumstances in which they are uttered, regardless of their syntactic position.

Of course, formally nothing prevents there being operators that bind the context parameter rather than the index parameter. Such operators would again bind the reference of indexicals and "disconnect" them from the context of utterance. Kaplan terms such operators "monsters", and he objects to introducing them into indexical semantics. His objection is based in one part on a quasi-empirical argument, and in another part on a conceptual argument. The quasi-empirical argument is that he could find no such operators in English. However, the range of data he was surveying did not come close to establishing the prohibition on monsters even as an empirical *hypothesis*, let alone an empirical *law*. The conceptual argument is therefore the more important, and it is based on Kaplan's semiotic thesis of direct reference. Kaplan's development of this thesis is not as clear and explicit as we would have liked, however, so the following account contains some reconstruction.

Kaplan's positive characterization of direct reference is not very systematic, and usually depends on metaphysical apparatus, such as structured propositions, which is contentious and arguably irrelevant to the task at hand. The nature of direct reference is more clearly glimpsed through its opposition with descriptive, or "sense-mediated", reference (p.485). Descriptive reference is a semantic notion. A descriptively referring term refers by virtue of a description being *true* of an individual in the domain of quantification. Logically, then, descriptive reference is dependent on quantification. (This fact is explicit in

9

Russell's approach to description, but the situation isn't essentially different with Strawson's presuppositional account, or with Carnap's and Montague's individual concept tactic.) Direct reference, by contrast, doesn't depend on the truth of a description, but is grounded in some kind of *contact* between the utterer of the referring term and the referent. This notion of contact can be thought of in causal or spatiotemporal terms, like Peirce's existential relation or Russell's logically proper names that refer in virtue of direct acquaintance (though acquaintance is not a purely causal or spatiotemporal concept).[16]

The contrast between the two modes of reference is then this. In order for the speaker to achieve descriptive reference, the referent doesn't need to be in any kind of contact with the speaker. It just needs to be in the domain of quantification and have (perhaps uniquely) the properties described in the denoting expressions. What quantification provides is an impersonal way of referring to objects, a way that doesn't, by itself, pick out any particular object. Picking out a particular object depends then only on the properties that the object has, that is on the predicates that it satisfies objectively. By "objectively" I here mean without any special reference to the speaker. The speaker is quite idle in descriptive reference. By contrast, direct reference picks out a referent regardless of its properties, based only on its contact with the speaker. There is no need here for the impersonal reference provided by the apparatus of quantification. This suggests that we think of quantification and direct reference as independent, in the sense that the objects referred to directly need not be in the domain of quantification. This sounds strange if we think of the domain as the collection of all things that exist. But there are independent reasons not to think in this way, namely the circumstance that the domain of a language will always exclude some things that exist (for example, the domain itself). Rather, I will here think of the domain as the collection of things that can be descriptively referred to in the language. I don't include things directly referred to (except if they happen also to be descriptively referred to) because quantification is idle in direct reference.

Conceiving things in this way shows clearly how the prohibition on monsters is derived from the direct reference thesis. If an indexical is bound by a quantificational operator over contexts, then its reference is no longer grounded in contact with its referent. The quantifier stands in between, as it were. For example, speaking at a time t, I can be said, in an obvious sense, to be in contact with t. The particular mode of contact between t and me is the one picked out the indexical 'now', so that 'now' refers to t. At time $t'_{\neq t}$ I no longer (or do not yet) stand in this particular mode of contact to t. Still, from within t' I can refer to t by quantifying over all times (e.g. by saying 'always'), regardless of my position with respect to them. This will no longer be reference by contact, however. Likewise, an indexical bound by an operator defined in terms of quantification over contexts would designate whatever it designates not by virtue of contact but through quantification. If Kaplan's direct reference thesis is read

---

16. The idea of contact is not explicit in Kaplan, but belongs to my reconstruction. See Kashtan (2020) for elaboration. See Hawthorne and Manley (2012) for a useful layout of the question of reference and acquaintance.

so as to say that indexicals *invariably* refer directly, then this is impossible, and we have to rule out context-quantifying operators.[17]

Does Kaplan's theory of indexicals belong to pure or to descriptive semiotics? It turns out that its two parts fare differently in each project: The double-index semantics is a descriptive success but a failure from the pure-semiotic perspective, and the other way around for the direct reference thesis. Consider, first, the descriptive demerits of the prohibition on monsters. Considered descriptively, the prohibition is an empirical prediction to the effect that context-shifting operators will never be observed in natural language. This prediction faces two problems. The first is that it has been falsified. In the past twenty years linguists have been gathering more and more evidence for phenomena for which the most natural explanation includes monstrous operators.[18] The second problem is that, even if the prediction had not been disconfirmed by Schlenker's data, we should be wary to begin with of empirical predictions made on the basis on conceptual or metaphysical considerations.[19] As noted above, Kaplan's approach to the empirical data was too cavalier for us to consider the prohibition on monsters as a proper empirical hypothesis.

By contrast, the double-index semantics has proved to be of enduring value for descriptive semantics, and indeed it is incorporated in one way or another into many formal semantic systems. However, if we look at it through the prism of pure semiotics, then it is easily seen to be trivial and uninteresting. We see this if we consider that, although the object-language is seemingly indexical, the metalanguage is perfectly standard. The Kaplanian object-language is nothing but a notational variant (slightly weakened) of a standard language with contexts in its domain of quantification.[20] From the pure-semiotic perspective, then, Kaplan's system does not embody or express any mode of designation different from the usual semantics for standard languages, and in particular, does not embody direct reference (though it can describe it).

What about the prohibition on monsters from the pure-semiotic perspective? Here, I contend, we have Kaplan's most valuable insight into the special mode of signification exhibited by indexicals. We have conceived of pure pragmatics as concerned with capturing direct, indexical, reference formally. Metasemantically, direct reference is reference in virtue of contact, and contact is severed by quantification (see previous subsection). A formal system capturing direct reference should therefore disallow any quantification that compromises the direct referentiality of indexicals. This is in essence what the prohibition on monsters says. In other words, from a pure-semiotic perspective, the prohibition is not a prediction, but an adequacy criterion on any formalism that purports to capture

---

17. Compare this explication of the prohibition on monsters with accounts that ground it in content-compositionality, e.g. (Rabern and Ball 2019).

18. Phillipe Schlenker has championed this approach; see Schlenker (2003) for the opening shot, (2017) for recent applications. Other work in this line includes Anand and Nevins (2004),(Major and Mayer 2019), and see references therein.

19. Kaplan himself is a little ambiguous as to whether the prohibition on monsters should be understood empirically.

20. See the appendix to Schlenker (2003) for work bridging the notational gap between indexed and standard systems.

(pure-semiotically) indexical reference.[21]

Stated in this way, it is clear that Kaplan's own double-index system fails to meet this adequacy condition. Kaplan himself refrained from defining context-shifting operators in fact, but they are straightforward to define in his formal system in principle. This is just a consequence of the circumstance, recently noted, that the double-index system is phrased in an essentially standard language, with standard quantification over contexts. It follows, or strongly appears to follow, that the prohibition on monsters cannot be captured (pure-semiotically) within a standard language. If we want to capture direct reference conceptually, we need to construct some novel kind of language. We are therefore in pursuit of a new expressive resource, one which significantly deviates from what we can find in standard languages.

2.2 Directive semantics

The first feature that the new expressive device has to capture is the "context-dependence" of indexicals, the fact that what they refer to depends on the circumstances (time, speaker, etc.) in which they are used. In Kaplan, the strategy is to reify constellations of circumstances of utterance into objects, the contexts, and state the reference of indexicals in terms of functions over such contexts. Where c is a variable ranging over contexts, the typical semantic value will be:

(5) $[\![{}'I'\!]\!]^c = speaker(c)$

This statement has an implicit initial universal quantifier binding c. Since the language in question (the metalanguage) is a standard language, its quantifiers nest freely, and monstrous operators can be defined for the object-language in a straightforward manner. The first-order quantifier is used here because this is the only way standard languages have of expressing the generality of a meaning entry (that it holds for all contexts), and, through the function expression, the dependence of the referent on the context. Our task now is to find an alternative way to express generality and context-sensitivity.

In his informal discussion, Kaplan distinguishes the *character* of an expression, which is the "*rule* for fixing the reference" in context, from its descriptive or intensional *content* (op.cit., p.515, emphasis added). The word 'rule' can be used in two senses: descriptively as a regularity that occurs, and prescriptively as a principle for directing behavior. Kaplan clearly means characters to be rules in the second sense, as principles that direct linguistic behavior. However, his formal system isn't equipped to capture rules in the prescriptive sense. Characters are modeled in the same way as contents, as functions from contexts to extensions.[22] Formally, the character of an indexical is the function abstracted from a syncategorematic semantic value statement of the kind of (5):

---

21. Making the prohibition on monsters definitive of indexicality is considered by Schlenker (2003, 32fn), who rejects it because it makes the prohibition empirically vacuous.

22. Actually Kaplan thinks of characters as functions from contexts to contents, but this is immaterial and somewhat misleading, as Lewis (1980, 90f) points out.

(6) $[\![{}'I'{}]\!] = \lambda c.speaker(c)$.

Functions, in the modern, extensional, sense, are essentially relations. In opposition to what its etymology suggests, a function in the modern sense is not something you *perform*. Function statements in effect do no more than describe relations in a notation convenient for various uses. Like relations, functions are defined on the domain of quantification. A practical rule, by contrast, is prescriptive: It doesn't describe anything, it *instructs* an agent to perform a certain action. A character is a rule instructing an agent to use language in a certain way.[23] By choosing to express characters as mathematical functions, Kaplan gives up on their prescriptive aspect at the outset, effectively building the dependence of indexical reference on standard quantification into the formal system.

The characters-as-functions issue, however, only highlights the problem, it is not its source. The prescriptive aspect of meaning was given up already in the syncategorematic (5), when we used a mathematical function to denote a context feature, and concomitantly, used the indicative mood for (5) itself. These are the two aspects of Kaplanian semantics that we need to revise if we want to redeem it as a pure-semiotic direct reference semantics. All sentences in standard languages, and in almost all formalized languages, are in the indicative mood. But indicatives are inherently incapable of expressing rules in the prescriptive sense; at best they can report regularities. The grammatical mood appropriate for expressing rules is the imperative mood, which is associated with directive rather declarative force. Accordingly, our strategy for capturing Kaplan's idea of rules for fixing reference will be to allow our semantic metatheory to form sentences in the imperative mood.[24]

The idea is to replace indicative semantic value entries like (5), that simply *describe* the referential patterns of indexicals, with directives that *instruct* linguistic agents how to fix the reference of the indexicals that they are using. Informally, such a rule will look something like this:

(7) When uttering 'I', refer by it to the utterer (yourself)!

There is more to say about the precise properties of these imperatives, but I will not do so here.[25] A speaker of the language, *situated* in a particular context, can choose to comply with the imperative and make their indexicals

---

23. Recall Strawson's (1950, 327) conception of the meaning of a term as "*general directions* for its use to refer to... particular objects...", and of a sentence as "*general directions* for its use in making true and false assertions" (original emphsis).

24. In naturally occurring discourse, imperative mood and directive force can diverge, but in theory we should still think of them as corresponding.

25. For example, such meaning directives don't resemble so much the commands or pleas that an individual might direct towards another individual (e.g. "pass me the salt, please", "get out of my way!"). They are more like standing normative injunctions, e.g. moral directives, the target of which is any individual who subscribes to the norm in question. In the case of moral directives, the addressee is any moral agent. In the case of meaning directives, it is any competent speaker of the language. One might then further ask, as one does with moral injunctions, what grounds their normative force. These issues lie beyond the scope of this paper.

refer as instructed to the objects they are in contact with. The metalanguage, by contrast, is *unsituated*, in contact with nothing. Nor is it necessary that it contain in its domain of quantification the objects that are in contact with the agents, nor even the agents themselves, and consequently not any reified context. In order for indexical reference to succeed, it is sufficient that the referent be in contact with the agent, and that the agent comply with the semantic directive as given in the metalanguage. Te metalanguage itself will be assumed not to be able to refer, generally, to agents, contexts, and context features.

There is, however, a sense in which the metalanguage does refer to these things. The informal meaning entry above contains the term 'the utterer', denoting a context-feature. In addition, it is formulated in the second-person (as imperatives are wont to do), so reference to the agent is implied as well. But, I submit, this kind of reference is not "concrete", ontologically committing reference. It receives a concrete referent only once it is complied with by a situated agent. Until then it is a floating or abstract kind of reference, or, as I will call it, *agent-pending*.

We formalize these ideas as follows. To the semantic metalanguage we add the symbol '@', called a *pseudo-variable*, which functions syntactically as a singular term that combines with function-like expressions to yield *pseudo-terms*, e.g. "*speaker*(@)", "*time*(@)". The pseudo-variable refers abstractly (in the sense given above) to contexts of utterance; the pseudo-terms refer abstractly to features of these contexts.[26] The pseudo-variable also functions as the marker of the imperative mood, or the carrier of directive force. A sentence containing @ is a *pseudo-sentence*, a directive for an agent to make a statement. Meaning statements such as (5) are replaced by entries such as the following:

(8) $\llbracket I \rrbracket = speaker(@)$

This whole sentence is to be understood as a directive to make the reference of the string "I" identical with the speaker. Similar entries can be given to the other indexicals. Note that this lexical entry, unlike the declarative (5), does not have an implicit initial universal quantifier over contexts, nor over '@', which is not really a variable at all. Since there are no quantifiers over contexts, such quantifiers cannot be nested within the right-hand sides of meaning entries.[27] It follows that context-shifting operators, monsters, are not definable in this semantics, in other words, that the prohibition on monsters is observed.

The prohibition on monsters was put forth above as an adequacy condition on an explication of the semiotic notion of direct or indexical reference. The

---

26. There is a subtlety here which I will gloss over, namely that on this semantics the agent-pending reference to a context never becomes proper reference, since the agent need not (and cannot) refer to the context they're in. This has some important consequences which, however, I will not explore here. See Kashtan (2018).

27. The pseudo-variable can be repeated. For example, the sentence:

(*) *speaker*(@) is in *place*(@), at time *time*(@).

is a directive to assert a sentence true just if the asserter is in the place of utterance at the time of utterance. But the pseudo-variable is not nested here, but used side by side.

condition being fulfilled by the directive semantics sketched here, we have the right to look at it as a pure-semiotically adequate direct reference semantics, or in other words, a successful system of *pure pragmatics*.

What has all this to do with the the idea of pragmatic mention, which according to §1 is the topic of this paper? The key, of course, is the notion of ontological commitment. In §1.2 we distinguished between material (semantic) and pragmatic mention in terms of the ontology they necessitate in their matrix language. Materially mentioning an expression involves referring to what the expression refers to, in the sense of having it in the domain of quantification. Mentioning an expression pragmatically does not involve this, though it does involve something else, a dependence on an external agent who takes "upon themselves" the ontological commitment. In other words, what we called abstract or agent-pending reference is our explication for §1's intuitive notion of pragmatic reference. On this way of parsing things, the reason Kaplan's double-index semantics goes wrong is that it relies on material rather than pragmatic mention.[28]

3 Directive semantics for formalized languages

3.1 Presemantic context sensitivity

The former section is about indexicals, a phenomenon from natural language. Whether the proposed directive semantics has any value for the study of natural language I don't know. Natural language served for us as a context of discovery, a bed of intuitions, and it might be that the notions developed, if considered with more theoretical rigor, will turn out not to correspond to anything real in natural language.[29] Our real object are formalized languages, the structure and properties of which are *stipulated*, rather than *given* empirically or phenomenologically. Formalized languages are in general developed in order to allow discourse which is unambiguous and logically perspicuous. One major hindrance to perspicuity is precisely context sensitivity, which makes interpretation depend on more than just linguistic form. Accordingly, indexicals are usually not to be found in formalized languages. How then are we to apply the ideas of the previous section?

In (1989, 559f), Kaplan distinguishes between "semantic" and "presemantic" context-sensitivity. Briefly, the distinction is between, on the one hand, context-sensitivity that is part of the language, as in the case of indexicals, where the

---

28. Note that what we have said about directive semantics doesn't exactly fit the notion of pragmatic mention as illustrated in example (4) of §1.2. This is because in that example, not only the pragmatically mentioned expression, but also the mentioning expression, is used by a situated agent, and not by an unsituated linguistic competence. That example was only included in order to bring pragmatic mention closer to intuition, and will not be handled in this paper. Indeed, it involves some important philosophical complications I don't yet have a good handle on.

29. That is, it might turn out that there are no proper indexicals in natural language, where a proper indexical is defined as directly referential.

meaning clause itself contains a relativization to context; and on the other, the more general fact that the meanings of strings don't belong to them intrinsically, but only as belonging to a language. The question which language to interpret a string as belonging to is answered by the circumstances of its use, so this is a further case of context-sensitivity. That the string 'I' refers now to me, now to you, is a semantic indeterminacy within English; that it may be interpreted now as the first-person pronoun (as in English), now as a non-contrastive coordinating conjunction (as in Polish), is a presemantic issue.

Kaplan's immediate goal in making the distinction is the claim that proper names, especially if Kripke's causal-chain picture of reference-fixing is accepted, belongs to pre-semantics, unlike indexicals. But we can exploit it for our own ends, as follows. Of course, every word of every language is subject to presemantic context sensitivity. The concept of context here is the same, and we can apply directive semantics also in the presemantic case. On this view, not only the meanings of indexicals, but the semantic value of every expression is made out to be a directive for an agent to use strings in a certain way - fixing reference or endowing with truth conditions. Nor is there reason not to proceed in exactly this way also in the case of formalized languages. Unlike their natural sistren, formalized languages are not (directly at least) the product of historical accident, but are due to explicit convention between people. This convention is what determines that, say, the string '∀' will express the universal quantifier, with the logical rules it is associated with. We can therefore think of this founding convention as the (presemantic) context of a formalized language (there is no need to associate a location and date with a convention). Given a convention C for a language L, every expression of L is interpreted according to C. In fact, it seems safe to identify a language L with the convention C that defines it (without deciding right now what kind of object a convention is).

What is the content of these new general meaning directives? Since formalized languages can be used to talk about anything at all, and since our new contexts are not assumed to contain any privileged points such as speaker and time, all a directive can do is instruct an agent to use an expression according to the prevailing convention. That is, the general form of a meaning directive is:

(9)   Endow the string s with the meaning dictated by the convention in place!

Actually, anyone using a string s in the presence of a convention for interpreting s will automatically be endowing it with the meaning dictated by that convention. Therefore the general form of a meaning directive can be given by the shorter formula:

(10)   use s!

Let's call imperatives of this form *use directives*. In the absence of any more definite stipulations about languages, this is the most precise that a completely general directive presemantics can do. It seems little, but as we will see, it is just enough to deliver some important results.[30]

---

30. Actually, if we limit ourselves to standard languages, a general directive presemantics

With these ideas in hand, I proceed to lay out the system of formalized languages I have in mind.

3.2 Standard languages

The formalization framework I will restrict myself to is that of classical extensional first-order predicate-quantifier languages (the standard languages). Many other kinds of language exist, but they can usually translated into standard languages by postulating special objects in the domain of quantification (e.g. possibilia for intensional languages; higher order objects for higher order languages, etc.), so the limitation to standard languages doesn't imply a severe limitation of expressive power. Although this kind of language is very well known, it will be useful to rehearse its definition, especially since we are interested in semiotics (the ways of linguistic designation) and not in logic (the patterns of valid deduction), and standard languages are most often defined with logic in mind. In particular, the languages we are interested in are *fully* interpreted, whereas in logic fully interpreted languages are only an auxiliary device.[31]

Let A be a set of symbol types called the *alphabet*. Symbols from A can be instantiated (tokened) and concatenated, i.e. placed one to the right of the other on a hypothetized row. Zero or more concatenated symbols form *a string*. The class of all strings P will be called *phonology*. Using certain conventions, we can associate various definable subsets of P with syntactic categories: variables, primitive predicates of arity n, well-formed formulas ($wffs$), well-formed formulas with n free variables (wff_n, for every n), etc. This is the *syntax* of formalized languages. Given the completeness theorem for recursively enumerable proof systems for first order logic, we can also define, in phonological terms, a relation $E(x,y)$ which holds between sentences x, y just in case x logically entails y (though see below). We assume phonology, syntax and logic to be the same across all languages.

A particular language is identified with an interpreted lexicon. A *lexicon* is a set of primitive predicates, individual constants and function symbols. The *interpretation*, or semantics, of a language L consists of a domain of quantification plus the definition of a binary relation of *satisfaction* between wffs and sequences of objects from the domain, along Tarskian lines. We can think of such a definition as an inductive definition with a base clause and with step clauses. The step clauses are the same across all languages and correspond to the logical terms; the base clause defines satisfaction for the lexicon, and is specific to a language. A string that is syntactically a wff and that is given satisfaction conditions by the semantics of L is a wff of L. Our notions of ontology and ideology (see 1.2) are easily captured - ontology is simply the domain, and ideology is the class of subsets of (Cartesian powers of) the domain definable by wffs of L.

---

can also say a great deal about syntactic structure and logical entailment. I won't go into this here, see Kashtan (2018).

31. In practice, fully interpreted languages may not be very useful, but they are important theoretically.

Technically, inductive definitions proper are not possible in standard languages. Instead, interpretations are defined explicitly, recapitulating the clauses of the informal inductive definition as closure conditions on sets. For this, one has to quantify over sets of a higher rank (or generally, objects of a higher order or level) than what the domain of the object-language contains. Hence, in particular, the language *in* which an interpretation is given never coincides with the language *for* which the interpretation is given, in other words there is no *semantic closure*. Consequently, our framework contains an unbounded "Tarskian hierarchy" of formalized languages, and in particular there is no ultimate or highest metalanguage. This framework, although it resembles it, is different from the more common model-theoretic semantic framework, in which the language of set theory is treated as an ultimate metalanguage and at the same time as a standard language.[32]

3.3 Directive semantics

We can easily locate the modes of use, formal mention, and material mention within this framework, and characterize their matrix languages. What it means to *use* an expression e of a language L_e in this framework is reasonably clear, as it is assumed that standard languages are *usable*.[33] The matrix language in this case is simply L_e itself.

Let L_{str} be the language that has the phonology P as its domain along with names for the phonemes of A and a function symbol for phoneme concatenation in its lexicon. In L_{str} we can form the names of all the strings and define all syntactic concepts. This is the string-theoretic language (or Tarski's "structural-descriptive" naming method), and it serves as the matrix language in all cases of *formal mention*.

In §1.2 I equated *material mention* of an expression e with reference to a pair $\langle s_e, L_e \rangle$, consisting of the the string and the language of e, respectively. It follows, first, that the matrix language M_e in the case of material mention extends the language of formal mention L_{str} (since it has to refer to s_e); and second, that has the language L_e in its domain, where L_e is identified with its semantic function.[34] By the impossibility of semantic closure, M_e will be distinct from, and higher in the hierarchy than, L_e.

The notions of use, formal mention and material mention find their cozy places in the framework as it stands. Accounting in a similar way for pragmatic mention requires complementing the framework with the device of use-directives,

---

32. One consequence of this is that we can't rely on the completeness theorem to get a syntactically defined logical entailment predicate, unless some sort of Kreiselian "squeezing argument" can be made. See (Smith 2011) for a brief explanation.

33. See footnote 5.

34. This second requirement can be weakened. If truth-conditions are sufficient for interpretation (as we will assume), then in order for a metalanguage to interpret an object-language, it is sufficient for it to have a truth predicate for the object-language (or a richer semantic predicate like satisfaction). This is a slightly weaker requirement than having the object-language in the domain of the metalanguage, but it doesn't amount to a significant difference, as far as I can see.

as indicated in §3.1. We define a language M_d, which extends L_{str} with the symbol '@', called a *use operator*.[35] Syntactically, the use operator combines with singular terms to form (atomic) *pseudo-wffs* (pwffs). pwffs then combine as wffs do: either with quantifiers and variables, or with connectives and other wffs or pwffs, and the result is always a pwff. If it has no free variables, a pwff is a pseudo-sentence, or psent. Thus syntactically pwffs function just like wffs. Semantically, however, pwffs are not interpreted. No sequence of objects satisfies a psent, but since the negation of a psent is itself a psent, we get that neither a psent nor its negation is true. This is because psents and pwffs are sentences in the imperative mood, and such sentences are not truth-apt at all. We therefore shouldn't think of them as truth-value gaps, anymore than chairs and tables and anything that is not truth-apt to begin with should be considered a truth-value gap. (Granted, I have spoken of "the negation of a psent", but since psents are not semantically interpreted, this is not a genuine negation.)

A special class of psents are those of the form $\ulcorner @\alpha \urcorner$, for α the constant name of a sentence. Such psents are *use directives* (see §3.1). They are the ones that are "passed on" to the pragmatics, that is, they instruct a prospective user of a formalized language to use the string α in the language that the agent happens to be using (presemantic context-sensitivity). The agent may then comply with the directive. psents that are not use directives are not candidates for compliance. The compliance of an agent with a directive $\ulcorner @\alpha \urcorner$ results in an instance of use of the string α in the language of the agent.

This can be symbolized for illustration. Compliance with a directive we write as square brackets around the directive; the relation of "resulting" from such compliance as a double arrow: $\Rightarrow$. An instance of compliance will be symbolized as here:

(11) $[@'Pa'] \Rightarrow Pa.$

Here the psent $@'Pa'$ is a directive to use the sentence $'Pa'$. Compliance with the directive (symbolized $[@'Pa']$) results ($\Rightarrow$) in an instance of *use* of the string 'Pa' (symbolized simply Pa). Note that this symbolization scheme is more illustrative than properly expressive. For one, the two sides of $\Rightarrow$ are interpreted as belonging to two different languages: the left-hand side belongs to the language M_d, and the right-hand side to the language of whatever agent chooses to comply with the directive, i.e. it belongs to an indeterminate language. As in the case of formal mention, we have a single matrix language M_d for all cases of *pragmatic mention* (in standard languages). But this language is not itself a standard language.

How do pwffs and psents interact with the logic? I'm not sure there is a natural answer, but for our purposes, as you will presently see, we need to stipulate that, although semantically unprocessed, pwffs and psents behave just as wffs and sentences do in logical inferences. For example, from the M_d psents:

---

35. We don't use the symbol in the same way as we did in §2.2, where it was called a pseudo-variable. In both cases it captures pragmatic mention, but in §2.2 only for indexical singular terms, and here generally.

(12) $Pa \rightarrow @'Qb'$,

(13) Pa,

I can use Modus Ponens to derive the use directive (still in M_d):

(14) $@'Qb'$.

Compliance with the directive yields, by analogy with (11):

(15) Qb.

The language of this last sentence is the language of the complying agent.

4   Application: Schema

Directive semantics, I think, can help solve certain problems in formal philosophy. In this paper we'll use it to solve a problem you might not have known you had - the problem of schemata. Schemata are used very often and very smoothly in the formal sciences, so my first task will be to convince you that there is a problem to begin with. The problem rises once we try to apply our methodological principle from §1.2, which instructs us to always be able to say which language we are using.

4.1   Peano arithmetic and formal mention

When formalized in a first-order language of arithmetic L_{PA}, the Peano-Dedekind induction axiom takes schematic form. All instances of the following schema are taken as axioms:

(16) $\phi(0) \wedge \forall n(\phi(n) \rightarrow \phi(n+1)) \rightarrow \forall n\phi(n)$.

Here $\ulcorner\phi(\alpha)\urcorner$, for α a singular term having at most 'n' as its free variable, is to be replaced by the result of replacing the free variable of a wff$_1$ of L_{PA} with α. For example, if ϕ is '$x + 1 = 1 + x$', then the following is an instance:

(17) $0 + 1 = 1 + 0 \wedge \forall n(n + 1 = 1 + n \rightarrow (n + 1) + 1 = 1 + (n + 1)) \rightarrow \forall n(n + 1 = 1 + n)$.

Since there are infinitely many suitable wffs, we have infinitely many such instances. This infinite set is the basis for proofs by induction in first-order Peano arithmetic. For example, we prove the sentence '$\forall n(n + 1 = 1 + n)$' by proving the two antecedents of the conditional in the schema instance (17), and then applying Modus Ponens to the instance itself. The schema guarantees that we have such an instance for every sentence in the language, and in particular for sentences of any complexity. There is no practical problem with this procedure, but it has not been formalized in any language, so we really don't know what is involved in it conceptually. The schema delineates an infinite collection of sentences of L_{PA}, any of which we can include at will in a proof. We need the schema because we can't directly *use* infinitely many sentences. But the

schema itself is not a sentence of L_{PA}, and it is incumbent upon us to say which language it *is* in.

If the instances of a schema are not *used*, then maybe they are *mentioned*. We attempt first to account for schemata using the notion of *formal mention*. In fact, the operation of instantiating a schema can be captured by a simple string-theoretic function, easily definable in L_{str}:

$$(18) \quad inst(x) = \underline{x[0]} \underline{\wedge \forall n(x[\underline{n}} \underline{\rightarrow} \underline{x[n+1])} \rightarrow \forall n(\underline{x[\underline{n}]}).$$

Here $x[y]$ is the replacement of the free variable of a wff$_1$ x by a singular term y (under the conditions stated below (16)); symbols are named by underlining and a concatenation of string names is a name for the concatenation of the named strings. The problem is that (18) is not a sentence, but a function. This highlights the fact that (16), by itself, is not the induction schema at all. In order to capture induction we need to look at the informal instructions above and below the schema. These instructions tell us to apply the function (18) to arbitrary wff$_1$s of L_{PA}, and to think of the results as axioms. But what does it mean to think of something as an axiom? Clearly it is not enough to *label* it as an axiom.[36]

In any case, we've made some ground, since we now know that the sentence that captures the work of the induction schema has the following logical form:

$$(19) \quad \forall x(wff_1(x) \rightarrow \Box(inst(x))).$$

This is the sentence from which we will want to derive the instance (17). Here $\Box$ is a placeholder for an expression, the precise identity of which we still have to discover, that has the effect of taking a string as an axiom. All the other expressions belong to L_{str}, so the language of the schema will be one that extends L_{str}, and the question is what else it includes. This will be answered when we discover what replaces $\Box$.

Notice that the language in question can't be L_{str} itself. No predicate of L_{str} can get us from $inst(\alpha)$, for any string α, to (17). (17) itself mentions numbers, which are not in the domain of L_{str}, so it is not a sentence of L_{str}. There is no problem to *mention* all of the axioms of L_{PA}, and indeed all of its theorems as well, in L_{str}. This is what we do when we talk *about* arithmetic, metamathematically. But in order to draw conclusions *in* arithmetic from the metamathematical results, we again have to somehow go from mentioning strings to using them, and this exceeds what L_{str} (or indeed L_{PA} itself, in its seat as a language of formal mention via coding) can provide.

4.2   Material mention and ZF

If the problem were only that of being able to refer both to strings and to numbers, an amalgam of L_{PA} and L_{str} might have done the trick. But we need

---

36. Labeling the instances as axioms is perfectly suitable for the purposes of proof theory, but not for us.

something more, namely to connect the right strings with the right numbers, and in general to be able to interpret the strings that we mention. But this is exactly how we characterized *material mention* in §1.2. Maybe we can formalize schemata using material mention?

In §3, material mention was equated with formal mention in the presence of a truth predicate. The truth predicate is what takes us from talking about strings to talking about numbers. On this view, the logical form of (16) is:

(20) $\forall x(wff_1(x) \rightarrow true(inst(x)))$.

The truth predicate entails all instances of the T-schema:

(21) $true(\ulcorner\phi\urcorner) \leftrightarrow \phi$.

Here 'ϕ' is replaced by a sentence and '$\ulcorner\phi\urcorner$' by a name of that sentence. From (20) together with (21) we can derive the desired instance (17).

There are two important problems with this approach. The first is that it relies on a schema. After all, our task is to give a general account of schemata, and an account that needs a schema in order to begin doesn't cut it. At best, we have managed to reduce all schemata to the T-schema – no mean feat, perhaps, but not a general solution to the problem. However, even if we decide to ignore this problem, in point of fact the truth predicate does not entail the instances of the T-schema after all, at least not as stated. The substitution condition I gave for (21) is in fact incorrect. It instructs us to instantiate a schema with the very same sentence mentioned on the left-hand side and used on the right-hand side. But this involves semantic closure, which is impossible. The proper substitution condition in Schema T is one that allows the sentences on the two sides of the biconditional in (21) to be distinct, and indeed to belong to different languages. The upshot of this for us is that (20), if it is to be applied to L_{PA}, cannot be formulated in L_{PA} itself, but in a stronger metalanguage M_{PA}. Consequently, it can't be used to derive the schema instance that we need, at least not directly, since the relation of logical entailment, as usually defined, holds only between sentences of the same language.

Maybe this is not such a dire problem. After all, there is an intuitive sense in which the two sentences replacing 'ϕ' in the T-schema, though they are in different languages, have the same content, are translations of each other. Maybe we can use some translation procedure to allow us to assert the schema instance (17), in L_{PA}, on the basis of the right-hand side of the corresponding instance of the T-schema, which is in M_{PA}. This suggestion is difficult to assess without a more explicit characterization of such a translation procedure, but intuitively it sounds as though it might work.

But this approach, though it might solve the problem in the case of arithmetic, would not constitute a general solution. Let's call a solution *regressive* if, in order to formalize a schema that belongs to a language L, uses a sentence in a more expressive language M. We should not be satisfied with a regressive solution to the problem of schema. For L_{PA}, positing a stronger metalanguage is not such a dire cost. L_{PA}, for all its virtues, is a rather weak language and we can imagine many metalanguages for it. The situation is different when we

consider set theory and its language. The standard set theory ZF, like Peano arithmetic, contains an infinite collection of axioms captured by the replacement axiom schema, and the use of this schema in proofs is comparable to the use of the induction axiom schema in arithmetic. But in this case we are hard-pressed to find a natural metalanguage that can materially mention set-theoretic expressions. Not that such metalanguages cannot and have not been proposed, but they have not achieved widespread acceptance, and in general we should prefer a non-regressive solution.[37]

4.3   Schemata and pragmatic mention

As the alert reader has surely already perceived, my discussion of the problem of schemata was tailored to lead us through the modes of the use-mention distinction. Schemata can't be explained by the category of *use*, since they encompass infinitely many sentences; they can't be captured by *formal mention*, since they involve interpreted sentences; and they can't be captured by *material mention*, because of "the ghost of the Tarski hierarchy". The notions of *pragmatic mention* and of the *use directive*, by contrast, offer a simple and intuitive formalization of our actual procedure with schemata. On this view, the logical form of the induction schema (16) is this:

(22) $\forall x(wff_1(x) \rightarrow @inst(x))$.

This is a sentence in M_d, the language of pragmatic mention, containing nothing but reference to strings and the use-operator '@'. Let 's' abbreviate our induction schema instance (17) (an L_{PA} sentence), let '$\overline{s}$' abbreviate a name of s (in L_{str}), and let '$\overline{u}$' abbreviate a name of the wff$_1$ of L_{PA} from which the instance (17) is generated. The following can be shown to hold:

(23) $wff_1(\overline{u}) \wedge \overline{s} = inst(\overline{u})$.

Since logic works in M_d as usual (see the end of §3.3), from (22) and (23) the following can be derived:

(24) $@\overline{s}$.

This is a use directive. Compliance with it is expressed as here (§3.3):

(25) $[@\overline{s}] \rightrightarrows s$.

Here s is to be read as an expression *in use*. Since by our setup, the language that the complying agent is using is L_{PA}, s is being used as a sentence in L_{PA}, not mentioned as a mere string. This overcomes the problems that beset the formal mention approach. Since M_D expresses use directives for all languages,

---

37. There is more to say here. For example, the so-called NBG class theory can be finitely axiomatized, and covers the same ground as ZF set theory. NBG is couched in a non-standard language, however, and has its own conceptual difficulties. There is a sense in which we do want (the language of) ZF to be a kind of an ultimate language, which it can't be if it is a standard language. My inclination is to view ZF not as a single standard language but as itself an unbounded hierarchy of languages, but I will not enlarge on that here.

this procedure is not regressive, so we have overcome the problems that beset the material mention approach. Deriving an instance from the formalized schema can be repeated for any wff$_1$ of the language, according to need. In other words, (22) expresses the *potential* use of any of the infinitely many induction axioms. In this way the problem of *actually* using infinitely many sentences is overcome as well. We thus declare the problem of the schema solved.

5 Concluding remarks

The use-mention distinction is fundamental for the philosophy of logic and language. In this paper I have tried to show that it is more intricate than usually believed, and in particular that it includes a mode of pragmatic mention. The theoretical basis for the notion was drawn from a critical analysis and a constructive alternative formalization of Kaplan's direct reference thesis for indexicals, in the form of a *directive semantics*. This semantics in turn was applied to formalized languages through Kaplan's notion of presemantic context-dependence. The main construction was the language M_d, that can issue "use directives", which instruct a situated user to *use* a certain expression in the language dictated by the context.

Finally, it was argued that the well-known and ostensibly harmless device of schema in the formal sciences is not well enough understood from a philosophical standpoint, and that its proper explication requires the notion of pragmatic mention. The notion of pragmatic mention and of a use directive are offered as the basic device of *pure pragmatics*, the philosophical study of language use.

Among the many questions that remain to be asked and, possibly, answered, one stands out in particular. Our guiding methodological principle from §1.2 instructs us to be able to state the language of each expression that we are using. Since we have spoken at some length *about* the language M_d, mentioning several of its expressions, we would like to know which language we have been using, and how to formalize it. I leave this question to another occasion.

References

Anand, Pranav, and Andrew Nevins. 2004. "Shifty operators in changing contexts." In *Semantics and linguistic theory,* 14:20–37.

Bar-Hillel, Yehoshua. 1954. "Indexical expressions." *Mind* 63 (251): 359–379.

Burgess, Alexis, and Brett Sherman. 2014. *Metasemantics: New essays on the foundations of meaning.* Oxford University Press, USA.

Carnap, Rudolf. 1942. *Introduction to Semantics.* Cambridge: Harvard University Press.

Enderton, Herbert B. 1972. *A Mathematical Introduction to Logic.* New York: Academic Press.

Hawthorne, John, and David Manley. 2012. *The Reference Book.* Oxford University Press.

Kaplan, David. 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals." In *Themes From Kaplan,* edited by Joseph Almog, John Perry, and Howard Wettstein, 481–563. Oxford University Press.

Kashtan, David. 2018. "An Observation about Truth, with Implications for Meaning and Language." PhD diss., Hebrew University of Jerusalem, Israel.

———. 2020. "How Can "I" Refer to Me?" In *The Architecture of Context Sensitivity,* edited by Tadeusz Ciecierski and PawełGrabarczyk. Springer.

Korta, Kepa, and John Perry. 2011. *Critical Pragmatics: An inquiry into reference and communication.* Cambridge University Press.

Kripke, Saul. 1975. "Outline of a Theory of Truth." *Journal of Philosophy* 72 (19): 690–716. https://doi.org/10.2307/2024634.

Lewis, David K. 1980. "Index, Context, and Content." In *Philosophy and Grammar,* edited by Stig Kanger and Sven Öhman, 79–100. Reidel.

Major, Travis, and Connor Mayer. 2019. "What indexical shift sounds like: Uyghur intonation and interpreting speech reports." In *Proceedings of NELS,* 49:255–264.

Rabern, Brian, and Derek Ball. 2019. "Monsters and the Theoretical Role of Context." *Philosophy and Phenomenological Research* 98 (2): 392–416.

Recanati, Francois. 2016. *Mental Files in Flux.* Oxford University Press.

Schlenker, Philippe. 2003. "A plea for monsters." *Linguistics and philosophy* 26 (1): 29–120.

———. 2017. "Super monsters I: Attitude and action role shift in sign language." *Semantics and Pragmatics* 10.

Searle, John. 1992. *The Rediscovery of the Mind.* MIT Press.

Searle, John R. 1983. *Intentionality: An essay in the philosophy of mind.* Cambridge University Press.

Smith, Peter. 2011. "Squeezing arguments." *Analysis* 71 (1): 22–30. ISSN: 00032638, 14678284. http://www.jstor.org/stable/41237272.

Strawson, P. F. 1950. "On Referring." *Mind* 59 (235): 320–344. https://doi.org/10.1093/mind/LIX.235.320.

Wettstein, Howard. 1986. "Has Semantics Rested on a Mistake?" *Journal of Philosophy* 83 (4): 185–209. https://doi.org/10.2307/2026531.

Yalcin, Seth. 2018. "Semantics as Model-Based Science." In *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics,* edited by Derek Ball and Brian Rabern, 334–360. Oxford University Press.