## Universities of Leeds, Sheffield and York
## http://eprints.whiterose.ac.uk/

This is an author produced version of a paper published in **Nous.**

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/3328/

# The price of inscrutability

(February 28, 2007)

*In our reasoning we depend on the stability of language, the fact that its signs do not arbitrarily change in meaning from moment to moment.*

(Campbell, 1994, p.82)

Some philosophers offer arguments contending that ordinary names such as "London" are radically indeterminate in reference. The conclusion of such arguments is that there is no fact of the matter whether "London" refers to a city in the south of England, or whether instead it refers to Sydney, Australia.

Some philosophers have even suggested that we *accept* the conclusion of these arguments. Such a position seems crazy to many; but what *exactly* goes wrong if one adopts such a view? This paper evaluates the theoretical costs incurred by one who endorses extreme inscrutability of reference (the 'inscrutabilist'). I show that there is one particular implication of extreme inscrutability which pushes the price of inscrutabilism too high. An extension of the classic 'permutation' arguments for extreme inscrutability allow us to establish what I dub 'extreme indexical inscrutability'. *This* result, I argue, unacceptably undermines the epistemology of inference.

The first half of the paper develops the background of permutation arguments for extreme inscrutability of reference and evaluates some initial attempts to make trouble for the inscrutabilist. Sections 1 and 2 describe the setting of the original permutation arguments for extreme inscrutability. Sections 3 and 4 survey four potential objections to extreme inscrutability of reference, including some recently raised in Vann McGee's excellent (2005a). Sections 5 sketches how the permutation arguments can be generalized to establish *extreme indexical inscrutability*; and shows how this contradicts a 'stability principle'—that our words do not arbitrarily change their reference from one moment to the next—which I claim plays a vital role in the epistemology of inference.

The second half of the paper develops in detail the case for thinking that language is stable in the relevant sense. In section 6, I use this distinction to call into question the epistemological relevance of validity of argument types; Kaplan's treatment of indexical validity partially resolves this worry, but there is a residual problem. In section 7, I argue that stability is exactly what is needed to bridge this final gap, and so secure the relevance of validity to good inferential practice. Section 8 responds to objections to this claim.

An appendix to the paper provides formal backing for the results cited in this paper, including a generalization of permutation arguments to the kind of rich setting required for a realistic semantics of natural language.[1] Extreme indexical inscrutability results can be proved within this setting.

The first half of the paper shows that the inscrutabilist is committed to extreme indexical inscrutability, which implies that language not determinately 'stable'. The second half of the

1

paper argues that good inference requires stability. The price of inscrutabilism, therefore, is to sever the connection between the validity of argument-forms and inferential practice: and this is too high a price to pay.

<div align="center">

# I

</div>

## 1   Inscrutability Arguments and Interpretationism

I take *positive inscrutability arguments* to have the following form:

1. SENTENTIAL CONSTRAINTS: The sole criterion of success for a semantic theory is that it assign the right semantic value to sentences: if two semantic theory both assign all the right semantic values to sentences, then it is indeterminate which one is correct.

2. OVERGENERATION: Multiple assignments of subsentential reference generate the right semantic values at the level of whole sentences.

3. *therefore* INSCRUTABILITY: It is indeterminate what subsentential expressions refer to.

Whether or not OVERGENERATION holds is a technical question, amenable to formal proof or disproof. SENTENTIAL CONSTRAINTS, on the other hand, is a distinctive and controversial thesis in the philosophy of linguistic representation (the discipline sometimes called 'foundational semantics', 'metasemantics', 'the theory of meaning' etc).

SENTENTIAL CONSTRAINTS has been endorsed in one form or another by many theorists, including David Lewis (1975), Donald Davidson (1974; 1977; 1979), Hilary Putnam (1980) and W. V. Quine (1960). The basic thought is a natural one: it is at the level of whole sentences that language most directly makes content with the states of affairs that it is its job to represent. We assertorically utter sentences to say that thus-and-such a state of affairs obtains: and this is manifest in a correlation between occasions of uttering whole sentences, and the states of the world on that occasion.[2]

The idea of Quine, Lewis, Davidson, Putnam and others is that from this correlation between sentences and states of the world one may extract information about what kind of semantic values sentences can have. Semantic theory as a whole is constrained to fit this sentential semantic data. This is the view encapsulated in SENTENTIAL CONSTRAINTS.

The theorists differ amongst themselves as to how the extraction of sentential semantic data is carried out; and over the form in which the data is to be presented. Some examples: the simplest view formulates the data simply as assignment of truth-values to sentences. Sentences are paired with the True just in case they are the ones that ordinary speakers assent to.[3] Equivalently, one might collect the sentences paired with 'the True' into one big set: what Lewis (1984) calls 'total theory'. SENTENTIAL CONSTRAINT then becomes the requirement that correct semantic theory make-true (so far as possible) this one big theory. This package is known as 'global descriptivism'.

More sophisticated alternatives include Davidson and Lewis's respective versions of 'radical interpretation'. For Davidson, the sentential data is to be formulated as the set of 'T-sentences' (e.g. 'schnee est weiss' is true iff snow is white). There is a complex story about how the radical interpreter fixes on a set of T-sentences in the course of finding an overall theory that best explains the behaviour of her subjects. Lewis (1975) offers yet another package. The sentential data is this time to be formulated as a pairing of propositions (sets of possible worlds) with

sentences, where the pairing is extracted from information about conventions of truthfulness and trust prevailing in a community. In all the cases mentioned, there is the distinctive second stage: semantic theory is constrained solely to match the sentential data, however it might be formulated.[4]

This style of foundational theory of linguistic representation ('metasemantic' theory) I call *interpretationism*. On all the versions of interpretationism currently in view, SENTENTIAL CONSTRAINTS will be sustained.

(We can break SENTENTIAL CONSTRAINTS down into two distinct theses:

1. SENTENTIAL DATA: The data-set constraining the selection of semantic theory concerns the semantic values of sentences—e.g. a set of T-sentences, or a pairing of sentences with propositions.

2. BEST=FIT: The sole criterion of success for a semantic theory is fitting this data: if two theories both optimally fit the data, it is indeterminate which is correct.

In response to the inscrutability arguments to be discussed below, many sympathetic to interpretationism look to avoid BEST=FIT. A famous example is Lewis (1983, 1984). Lewis describes semantic theories as more or less *eligible* to the extent that the semantic values they assign to words are *natural* (so the thought is that all else equal, an interpretation that assigns *green* to a word, is more eligible than one that assigns 'grue' to that same word.) Lewis then argues within the context of a broadly interpretationist metasemantics, that the meaning-giving interpretation, as well as getting the semantic values of whole sentences optimally right, also needs to be optimally *eligible*. These complications to the interpretationist picture are not our topic here.)

Given SENTENTIAL CONSTRAINTS, to argue for the extreme inscrutability, it suffices to show that given a 'sensible looking' semantic theory for English, we can construct another one that:

1. Assigns the same semantic values to whole sentences.

2. Embeds arbitrarily crazy scheme of reference to singular terms

The famous 'permutation arguments' show us that such overgeneration results are indeed available.

## 2   Permutation arguments

The key idea of permutation arguments is that twisted assignments of extensions (referents) to constants are compensated by equally twisted assignments of extensions to predicates. Overall, the twists 'cancel out' to deliver the usual result at the level of sentences.[5] For a toy example of this, let us use the phrase 'the image of x' to pick out *x* whenever *x* is anything other than Billy or the Taj Mahal; to pick out Billy if *x* is the Taj Mahal; and to pick out the Taj Mahal if *x* is Billy. We can describe our twisted reference scheme as follows: *N* twist-refers to *x* iff *N* standardly-refers to *y* and *x* is the image of *y*. For the similarly twisted assignment of extensions to predicates, take any atomic predicate *P*. Let *P* twist-apply to *x* iff *P* standardly applies to some *y*, such that *x* is the image of *y*. The twists cancel out—the distribution of truth-values to sentences is the same on both interpretations. For example, "Billy is running" is true iff the referent of "Billy" falls under the extension of "runs". Now, the twist-referent of "Billy" is the Taj Mahal; but the twist-extension of "runs" includes all the images of running things. Since

Billy runs, and the Taj Mahal is the image of Billy, the Taj Mahal falls under the twist-extension of "runs". The sentence comes out true, just as it does under the standard interpretation. By induction, we are able to show that this generalizes to arbitrary truth-functional compounds.

The above sketch gives the root idea of permutation arguments: twists in one area of subsentential content 'cancelling out' others. To provide an inscrutability argument for English, however, we would need to argue that such a result goes through for a semantic theory adequate for the richness and complexity of syntactic structure we find in that language. Further, for full generality, we should show not only that permutation arguments preserve *truth values* but *whatever* entities are assigned as semantic values of sentences. The fullest presentation of these results of which I am aware[6] still only proves the result for a framework that is expressively limited in comparison to the multiply intensional type theory used by linguists in the Montague tradition. In an appendix to this paper, I provide a more general result: I show that the permutation results hold within the 'general semantics' framework of Lewis (1970).

# 3   The costs of extreme inscrutability of reference

Interpretationism is an attractive theory of how facts about linguistic representation can be fixed by facts about patterns of assent and dissent (facts that are do not themselves have to be characterised in semantic terms). In the form presented above, interpretationism is a very clean and simple theory. Yet it apparently implies extreme inscrutability of reference: there is no fact of the matter about whether "Londres" refers to London or to Sydney.

One reaction at this point would be to start tweaking the content of interpretationism: dropping the idea that the *sole* constraint on the success of semantic theory is that it 'fit data' about the semantic values of whole sentences (what we earlier called BEST=FIT). Such additions increase the complexity of interpretationism: for example, on Lewis's preferred version, mentioned earlier, we will need a (non-semantic) account of what it takes for macroscopic properties to be *eligible* or *natural*; we also require some explication of how to weigh up considerations of fit against considerations of eligibility.

I do not wish to *rule out* the tenability of some such tweaked version of interpretationism, especially if the impression that the additional constraints are *ad hoc* monster-barring moves can be ameliorated. But getting the tweaks right is no easy task. (I argue in (2006a) that Lewis's response fails unless we accept a radically 'emergent' ontology of fundamental properties.)

To start altering the theory may be an overreaction. There is much to gain if we could 'bite the bullet' and simply *accept* that reference is inscrutable: we gain a clean and simple theory of linguistic representation, and at what cost? Certainly some *counterintuitive* results. But (one may think) philosophers have learned to live with worse.

As already mentioned, Davidson (1977, 1979) is one such figure. On his view, commitment to extreme inscrutability is an *insight* of interpretationism, not a problem for it. The central purpose of the remainder of this paper is to examine whether this is a tenable proposal. We shall be looking at the *costs* of inscrutability of reference. In the end, I shall argue, inscrutability of reference poses unsurmountable problems for our theory of inferential practice. To emphasise: it is not my aim here to show where the argument for inscrutability goes wrong, if it does: it is to determine whether or not we are *obliged* to find some flaw in it.

Before turning to detailed considerations, however, we should distinguish between two forms of inscrutability thesis one might argue against in this way. For example, consider the English expression "Rabbit". Arguments for *moderate inscrutability* aim to establish that there is no fact of the matter which of a range of entities this word applies to. Famously, Quine's "argument from below" (1950, ch.2) aimed to establish that there is no fact of the matter whether 'rab-

bit' applies to rabbits, instantaneous rabbit-stages, or undetached rabbit-parts. McGee (2005a) argues that Unger's (1980) 'problem of the many' should be seen as establishing a similarly moderate form of inscrutability: there is no fact of the matter which entity the name "Kilimanjaro" picks out, though whatever it picks out will be a mountain-like agglomeration of rock in the vicinity of that mountain. Some think that the vagueness of much of natural language should be analyzed as consisting in a (moderate) form of inscrutability.

Contrasted with this putative moderate inscrutability, the form of inscrutability that the permutation argument targets is *extreme*. An extreme inscrutability thesis might say that there is *no fact of the matter whatsoever* which entities 'rabbit' applies to, or what 'Kilimanjaro' picks out. This extreme inscrutability thesis holds that Paris and Sydney, and small furry creatures orbiting distant stars, have as much claim as any city in the south of England to be the referent of "London". It is not determinately the case that 'London' doesn't refer to *them*. The permutation arguments accepted by Davidson and others purport to establish this extreme form of inscrutability.

My focus in this paper will be on extreme inscrutability results, and the costs of accepting that reference is inscrutability in the extreme way. But I will also consider whether the considerations I bring against extreme inscrutability are also reasons for rejecting moderate inscrutability. On the one hand, a successful argument for the *determinacy* of reference (against moderate or extreme inscrutability) would be a surprising and significant result. On the other hand, an argument that impacts *only* on extreme inscrutability, while leaving moderate inscrutability untouched, has more of a claim to be identifying the distinctive costs of the interpretationist metasemantic setting. Either way, therefore, it is worth considering the scope of the arguments we consider.[7]

# 4   Objections to inscrutability

In this paper, I will be largely focussed on investigating one particular cost of extreme inscrutability: its damaging interaction with the idea of good token inferences. But one might suppose that there were other, quicker objections to be had. So I begin by briefly mentioning some potential candidates, to flag up what I take to be problematic about them.[8]

**The incredulous stare**   The thesis under consideration is quite extraordinary. Hold up a red ball in front of your face. Say aloud "that red ball is shiny". Accompany the words by jabbing your fingers into the ball to remove any reasonable way of mistaking your intent. What is being claimed is that nevertheless, the words "that red ball" might just as well refer to the Taj Mahal as to the red, shiny ball in front of you. Isn't that just unbelievable? Moreover, the claims denied seem obvious truisms: it denies 'disquotational' reference principles such as ' "London" refers to London'.

For languages in general, and in particular for the reference of one's own language, arguments for extreme inscrutability seem to try to deny the non-negotiable. Lewis (1983, p.46) describes the principle that "Our language does have a fairly determinate interpretation" as "a Moorean fact"—one which we have more reason to believe, than we have justification for believing any premises that might undermine it. For these, and other reasons, one might think that the extreme inscrutability thesis is literally incredible.

Notice that the incredulity here is directed at *extreme* inscrutability theses. Moderate inscrutability—such as that purportedly generated in the problem of the many, or in obscure mathematical cases—does not, I think, generate quite such intuitive repulsion. Some, indeed, have felt something the purported limited inscrutability of 'rabbit' for

which Quine argues is incredible; elsewhere (2006b) I argue that this is an artefact of the way the Quine's argument is typically presented, and that the conclusion *should not* considered all that surprising. The objection by incredulous stare, therefore, is particular to the case of extreme inscrutability.

**Underminingness** One might suspect that accepting extreme inscrutability undermines the very argument we give for it. If the theorist's language is itself inscrutable in reference, it may be thought mysterious how we conceive of and discriminate between various rival interpretations of our language. In particular, are we not supposing that there is an 'intended' interpretation of the language, from which our 'deviant' interpretations are parasitically constructed? If so, doesn't the conclusion of the inscrutability argument undermine the argument given for it?

Again, we can ask the same question about moderate inscrutability theses: are they damagingly undermining? Of course, whether there's a *prima facie* case to answer depends on what exactly the argument is for this moderate inscrutability, that might be 'undermined' by the conclusion. (The most plausible case that I know of is that of *indeterminacy of quantification* arising from the Loweinheim-Skolem arguments as deployed by e.g. Putnam (1980). For to give an argument that one's unrestricted quantifiers are inscrutable in range between $D_1$ and $D_2$, it looks like one has to allow oneself the resources to quantify over the union of these sets! And the conclusion of the argument is that it is not determinately the case that one can do this.)

**Semantic beliefs** A traditional view in philosophy is that linguistic competence is a matter of (something like) *knowledge of the meaning of expressions*.[9] Call these *cognitive* approaches to understanding. It might be thought that extreme inscrutability (to its discredit) rules out a cognitive account of understanding.

The kind of cognitive approach to understanding that may be in tension with extreme inscrutability is one where we have to have knowledge, not only of what *sentences* express, but also what *words within sentences* express.[10] On this 'strong' cognitive conception of understanding, in understanding a sentence containing the name $N$ we must at least believe something of the form:

>   $N$ refers to $o$

or perhaps, for appropriate descriptive content $F$:

>   $N$ refers to something which is $F$

Extreme inscrutability says that there is never any fact of the matter about whether such claims hold good. *Prima facie*, the upshot will be that one's understanding will consist in part in beliefs whose contents are indeterminate. And this seems to lead the inscrutabilist into Moorean paradoxical situations: she holds $p$ to be indeterminate, but nevertheless (as part of her linguistic competence) believe that $p$. This seems to disclose a form of irrationality in the inscrutabilist stance.

Does the moderate inscrutabilist face parallel worries? If we insist that linguistic competence in $N$ is partially constituted by the semantic belief that $N$ *refers to o* then they are in the same situation as the extreme inscrutabilist: being committed to beliefs which are indeterminate in content. Under this assumption, moderate and extreme inscrutability stand and fall together. However, under a slightly more relaxed version of the strong

6

cognitive approach the understanding, the semantic belief that in part underlies our understanding of 'Kilimanjaro is big' can be something like: *'Kilimanjaro' refers to some mountain-like agglomeration of rock located in Southern Africa, larger than any other distinct mountain-like agglomerations of rock in that continent.* On the sort of moderate inscrutability generate by 'problem of the many' considerations, the descriptive content here invoked may determinately apply to *all* of the candidate referents for 'Kilimanajaro' (in the extreme inscrutability case, of course, any such predicate would have to apply to absolutely every object!). So, absent arguments that the strong cognitive conception has to take the more severe form, there is no *general* tension between moderate inscrutability theses and the strong cognitive conception of understanding.

**Inscrutability of rigidity and existence**  Vann McGee (2005a) has recently put forward a range of considerations intended to show that the permutation arguments for inscrutability break down when we consider more than just the expressively impoverished setting of extensional first order logic. An appropriate rejoinder is to show that the arguments *do* go through in as rich a setting as one could want. This is done in the appendix to this paper.

Though I do not think that McGee's arguments provide formal obstacles to the inscrutabilist, I think they show something really interesting about the *commitments* of the inscrutabilist.

McGee considers permutation arguments in the setting of quantified modal logic (like him I assume familiarity with standard possible worlds semantics for such systems). Now, it is commonly (though not universally) held that what exists in one world can fail to exist in another. Suppose that to be the case. When deploying one's permutation argument one can either (a) look at the domain of *possibilia* as a whole, and let one's permuted interpretation be based on the a permutation of possibilia irrespective of which worlds they exist at; or (b) take the domain of each world in turn, and permute the extensions of expressions at world *w where the each such permutation is restricted to objects that exist within its respective world*. Either way, McGee thinks there is trouble.

On strategy (a), suppose that the 'overall' permutation maps possibilium *a* to possibilium *b*, and that on the intended interpretation of our language the rigid designator 'Billy' is a name for *a*. Then (absent necessary connections between *a* and *b*) there will be a world *w* where *a* exists and *b* does not. On the intended interpretation, 'Billy' denotes *a* at *w*; and consequently on the permuted interpretation, 'Billy' must denote *b* at that world. But *b* doesn't exist even exist in *w*!

To dramatize this, suppose that *w* is the *actual* world, and 'Billy' a name in English. Then the permuted interpretation has it that 'Billy' names something that doesn't actually exist; and absent special pleading, the inscrutabilist has to say that there is no fact of the matter whether or not our ordinary proper names designate non-existent entities.

On strategy (b), we are forced to choose different permutations at different worlds. On this strategy the *intension* of 'Billy' will be a function mapping a world *w* to the image of *a* under the *permutation chosen for w*. Suppose *a* is mapped to *b* by the chosen permutation for the actual world. Then (on the assumption that there is no necessary connection between *a* and *b*) take a world *w* where *a* exists but *b* does not. The chosen world-restricted permutation for *w* obviously *can't* send *a* to *b*. So it must send *a* to something else, *c* say. So the two permutations disagree on what *a* is mapped to. Again, suppose 'Billy' names *a* according to the intended interpretation. On the permuted interpretation, the extension of 'Billy' at the actual world will be *b*, and the extension of 'Billy' at *w* will be the distinct object *c*. Thus, absent necessary connections between individuals, any "permuted alternative" to the sensible interpretation *the ordinary proper name 'Billy' will turn out*

*to be non-rigid.* If such deviant interpretations cannot be excluded by interpretationist lights, we get not only extreme inscrutability, but also inscrutability over whether or not ordinary proper names are rigid designators.

McGee's arguments seem particular to the case of extreme inscrutability: there is no obvious way to construct parallel concerns for the moderate inscrutabilist.

We have just seen a number of challenges to inscrutability, and to extreme inscrutability in particular. Notice that, in accordance with the spirit of this paper, none attempts to describe *what goes wrong* with the permutation argument.[11]

I am dissatisfied by each of the above putative objections to the inscrutabilist. I do not have space here to discuss each in detail, I will briefly mention the shape of my concern in each case.[12].

In philosophy, there is a track record of views that generate incredulous stares: David Lewis's extreme modal realism; Timothy Williamson's epistemicist theory of vagueness; Graham Priest's defence of true contradictions, Peter Van Inwagen's denial of the existence of medium sized dry goods. But in each case, their opponents seek objections *beyond* the incredulous stare, and with good reason.

If one's *only* grounds for complaint against a theory is that it is literally incredible, then one is in a dialectically weak position when faced with someone who, in all seriousness, advocates the position. Such an inscrutabilist may fairly object that the psychological limitations exhibited in *your* finding the position literally unbelievable should be of no concern to them. (And, recall, the 'extreme inscrutabilist' is no mere fiction: Donald Davidson accepts this position.)

Perhaps, though, the emphasis should not be on the 'literal unbelievability' of extreme inscrutability. Rather, we should focus on its *counterintuitiveness* or 'Moorean' status. Even the convinced inscrutabilist may admit that it is some cost to their theory that they go against common sense in these ways. But with what right do we suppose that the *discredit* this earns the inscrutabilist position is not outweighed by the *benefits* that the position makes available? I can only speak for myself, but the interpretationist promise of a reductive account of the representational properties of language would be, for me, a big enough benefit to compensate for some departures from pre-theoretic opinion *were there to be no independent problems with the view, and were I able to bring myself to believe in it*.

Were I to endorse inscrutabilism, I would seek to ameliorate the embarrassment of disagreeing with pre-theoretic opinion by attempting to 'explain away' intuitions against inscrutabilism. I might, for example, point to the fact that one can introduce *into the object language* devices that obey the disquotational rules for 'truth' and 'reference' respectively. I would carefully distinguish these purely disquotational notions from the heavyweight world-word relations that the interpretationist seeks to give an account of. Accepting extreme inscrutability for the representational notion of 'reference' is not to debar oneself from talking-the-talk of disquotational reference.[13]

The idea that adopting the thesis of inscrutabilism renders problematic the arguments given for accepting it is intriguing. AS already mentioned, it is extremely sensitive to the exact nature of the arguments for inscrutabilism given: with the putative moderate inscrutability of unrestricted quantification being a promising case. The question is: do the permutation arguments for extreme inscrutability have such a form? Admittedly, permutation arguments for inscrutability are often presented by talking about an 'intended' interpretation, and derivatively constructing deviant interpretations. That naturally leads to the thought that *absent* an 'intended' interpretation they will be ill-formed. However, the appearances mislead. We should simply note that we never need to pick out individual interpretations in giving the argument— not even an 'intended one'. Rather, we just argue for the *conditional* claim that, for all *x*, if *x*

is an interpretation-function which renders true a set of sentences $T$, then a permuted variant of $x$ also renders this true.[14] We then need only the principle that there is *some* interpretation which gets the sentential data right, to conclude to extreme inscrutability.[15] I conclude that these considerations fail to cause trouble for the inscrutabilist.

The putative tension between strong cognitive accounts of understanding and extreme inscrutability is a promising route. Indeed, I think that, aside from the approach explored at length here, this is the most promising way to argue against extreme inscrutability. It is important to note that the claim that linguistic competence in part consists of something like beliefs about subsentential meaning is itself highly controversial. So no *knock down* blow can be expected here. Elsewhere, I have argued that the inscrutabilist can resist the objection directly: by considering carefully potential *correlative* indeterminacies between the content of thought and the content of language, we are able to resist the argument that the semantic beliefs needed for the strong cognitive conception of understanding are indeterminate.

That leaves us with the McGee-inspired costs: the result that (depending on how we extend the permutation arguments) we are faced either by what we might call the inscrutability of rigidity, or the inscrutability of existence. Let me immediately note that this case is predicated on the assumption that the domains of worlds vary: that things go in and out of existence from world to world. This is certainly deniable.[16] Setting that line of resistance aside, notice that the considerations mentioned at best show *that the inscrutabilist is committed to one or other of the inscrutability of existence or the inscrutability of rigidity*. They do not show *what is wrong with being committed to one or other of these results*. Extreme inscrutability is already highly counterintuitive; what we have been shown is other, similarly counterintuitive consequences of the position. But again, is the alleged badness of these commitments of the inscrutabilist position simply that they generate an incredulous stare (in which case, we've made no progress, since extreme inscrutability alone does that!). Or are there more detailed considerations to be found against them?

It is true that the rigidity of proper names is philosophical orthodoxy, and it is natural to hope to find, in the arguments that established this orthodoxy, grounds for objecting to the inscrutabilist who holds that it is indeterminate whether proper names are rigid.[17] But the evidence for this claim, I take it, largely consists of data about the truth-conditions of various sentences (such as 'Aristotle could have died at birth'). But of course *the (non-rigid) permuted interpretations match the 'intended' one at the level of truth-conditions of whole sentences*. So there simply isn't going to be any data of this kind that tells against non-rigidity in this peculiar form.[18]

What about the inscrutability of existence? Is there something untenable about the thesis that one refers to (contingently) non-existent things? I do not see where this comes from. Notice that it cannot be a problem over words standing in relations to non-existent relations *as such* that is at issue: for on the kind of modal semantic framework that is being assumed, the semantic values (intensions) of predicates such as 'is self-identical' will be sets that contain mere possiblia within their transitive closure. So semantic relations already relate words to entities of provenance at least as dubious as possibilia themselves. Why should the semantic relation of *reference* be any different?

But once the ontology worry about relating-to-the-ontologically-dubious are set to one side, it is hard to see what the remaining worry is. Consider a temporal analogy. We can refer to *Beethoven*, even though he does not exist at the present time. Likewise, the current extension of 'famous' contains presently non-existent things, for Beethoven is currently famous. Such examples show, I take it, that some words refer relative to a time, to things that fail to exist at that time. Why should the relativization to worlds be any different?[19]

(Quite independently of the current considerations, some philosophers have argued that reference to mere possibilia can be achieved.[20] Salmon (1998) invites us to consider the merely possible person who would have developed from *this* egg and *this sperm*, had they fused and developed. He argues that (granted certain essentialist assumptions) we introduce a name for this merely possible person. Far from being impossible, reference to the non-existent is arguably *actual*.)

The working assumption shall therefore be that none of the above provide cogent arguments against the extreme inscrutabilist. I therefore turn to considerations that do seem to me to promise to show why extreme inscrutabilism is untenable.

# 5   Inscrutability, Context and Validity

**Extreme indexical inscrutability**

What is required of a semantic theory by SENTENTIAL CONSTRAINTS, is simply that it assign the right semantic values to sentences. And we have already seen a recipe for constructing semantic theories embedding arbitrarily permuted reference schemes, that are sententially equivalent to 'intuitively correct' semantic theory.

However, we need not stop here. Given SENTENTIAL CONSTRAINTS, we have all the resources we need to prove an even more general inscrutability result than the extreme inscrutability theses previously mentioned: *extreme indexical inscrutability*. This says that not only is there no fact of the matter as to whether 'London' refers to London or to Paris, but there is no fact of the matter whether or not it *changes* from referring to London, to referring to Paris, from one moment to the next.

One might wonder why indexical inscrutability doesn't follow *immediately* from simple inscrutability. There is some *prima facie* appeal to the following argument:

1. It is indeterminate whether $e$ denotes $o$ in context $c$;

2. It is indeterminate whether $e$ denotes $o$ in context $c'$;

3. *Therefore* It is indeterminate whether ( $e$ denotes $o$ in context $c$ and $e$ denotes $o$ in context $c'$)

Since we can run this argument for *any* object $o$, it looks like this form of argument can establish the inscrutability of co-reference across contexts, directly from the standard inscrutability of reference.

The form of inference, however, is fallacious: the logic of 'determinately' simply doesn't work that way. Here is an argument of similar form:

1. It is indeterminate whether 'Billy' denotes the Taj Mahal;

2. It is indeterminate whether 'Billy' denotes the Eiffel Tower;

3. *Therefore* It is indeterminate whether ('Billy' denotes the Taj Mahal and 'Billy' denotes the Eiffel Tower)

The extreme inscrutabilist again grants both premises. But of course they reject the conclusion: they agree with orthodoxy that *determinately* a name picks out just one thing. This illustrates that indeterminacies can be *coordinated* in a way that blocks the form of inference that might

get us quickly from extreme inscrutability to extreme indexical inscrutability. So more work is needed.[21]

The detailed case for extreme indexical inscrutability is built from the permutation arguments previously sketched, and in broad outlines is as follows.[22] First of all, we require a semantic theory that *in principle* allows words to vary in reference from one context to the next. But we need this in any case, since some words ('I' and 'you' and 'them', etc) do display such features. The standard Montagovian treatment of the phenomenon is to let the semantic value of a lexical item be a function from contexts to intensions. Thus, 'I' will be assigned a semantic value that takes contexts where Billy utters the word to the intension that is a constant function from worlds to Billy; and which takes contexts where Sally utters the word to the intension that is a constant function from worlds to Sally.

Suppose we take two 'permuted' semantic theories constructed in pursuance of the original inscrutability result, $\theta_1$ and $\theta_2$. These will each assign to expressions functions from contexts to intensions, which embed arbitrary crazy actual-world-reference schemes. In particular, suppose that $\theta_1$ represents 'Billy' as referring to the Taj Mahal (no matter what the context is) and $\theta_2$ represents 'Billy' as referring to the Great Wall of China (no matter what the context is).

Take any two contexts (say, the context in which I write these words, and the context in which you read them). The idea is then that we 'cut and shunt' these two interpretations to construct a *third one*, $\Theta$, that will represent 'Billy' as referring to the Taj Mahal in the former context, and as referring to the Great Wall of China in the latter. In short, unlike $\theta_1$ and $\theta_2$, it represents Billy as an indexical expression, and one that shifts its reference over the (arbitrarily chosen) context-change.

How do we do this? Simply by considering the set of all possible contexts, $C$. This contains, in particular, the context $c_1$ in which I am writing, and the context $c_2$ in which you are reading. It will suffice to say what intensions $\Theta$ assigns to words at each point in $C$. So divide $C$ exhaustively into two subsets, $C_1$, containing $c_1$, and $C_2$, containing $c_2$. Then say that the intension assigned by $\Theta$ to a word in a context $c$ in $C_1$ are exactly those assigned to it in $c$ by $\theta_1$; and the intension assigned by $\Theta$ to a word in a context $c$ in $C_2$ are exactly those assigned to it in $c$ by $\theta_2$. One can quickly check that this recipe means that 'Billy' will behave (according to $\Theta$) exactly in the indexical way described above.

To argue from SENTENTIAL CONSTRAINTS that $\Theta$ is a semantic theory which isn't determinately *incorrect*, we need to make sure that at $\Theta$ fits with the all the sentential data the interpretationist can lay down. We secure this result because of the way $\Theta$ was constructed out of two interpretations which (*ex hypothesi*) already meet this condition. No matter which context we choose, either the semantic content assigned by $\Theta$ to all of the lexical items will match that assigned by $\theta_1$ or the semantic content assigned by $\Theta$ to all of the lexical items will match that assigned by $\theta_2$. By the compositionality of language, the intension of the sentence (in that context) is determined by the intension of its lexical parts (in that context). So in any context, $\Theta$ will assign to sentences intensions that match those assigned by $\theta_1$ or by $\theta_2$. But we are supposing that at the level of sentences, both these semantic theories match each other and fit with all the data the interpretationist can lay down. Hence the compound theory is *sententially equivalent* to $\theta_1$ and $\theta_2$: the interpretationist cannot declare it determinately incorrect.[23]

The upshot is that for any name *n*, and any pair of contexts $c, c'$, a compound theory fitting the interpretationist's sentential data can be constructed, so that the reference assigned to *n* changes from $c$ to $c'$. And, by interpretationist lights, there is nothing to make *this compound theory* determinately 'unintended'. Thus, *extreme indexical inscrutability* follows: there is never determinately the case, from one context to the next, that the reference of a name remains the same.

Extreme indexical inscrutability requires only a minor extension of the original permutation

arguments for extreme inscrutability of reference. If an interpretationist subscribes to extreme inscrutability on these grounds, she is in no position to resist my argument for extreme indexical inscrutability. As we shall see, the price extracted by extreme indexical inscrutability is high.

## Good inference and context

Consider the following schematic claim:

> STABILITY
> For the set of contexts $C$ (with at least two elements), and the non-null set of expressions $E$, the extension of each expression in $E$ is the same at each context in $C$.

Extreme inscrutability, I have argued, will generalize to extreme indexical inscrutability. And it should be clear that extreme indexical inscrutability would entail that no instance of STABILITY is better than *indeterminate* in status.

The second half of this paper is concerned to measure what costs are incurred in giving up all instances of STABILITY. We shall focus on the standing of certain *token* inferences which are of a logically valid *type*: arguments such as 'Annie is female; Jen is female; so Annie and Jen are female'. Such arguments typically enjoy *logical* good standing: they have a distinctive (and arguably more impressive) epistemic status than arguments that are of an inductive form (all observed students are female; so all students are female); or which are backed only by *entailments* rather than logical validities (this is water; so this is $H_2O$). I will argue that such in spelling out what it is for a token argument to be in logical good standing, it will be a requirement that the expressions involved in the argument be stable across the contexts involved in the token argument. Thus, if all instances of stability are given up, no token argument will ever be (determinately) in logical good standing.

We can turn this into an explicit argument against extreme inscrutability, which forms the 'master argument' of this paper. Suppose that $E(z)$ is the set of type-expressions tokened in (token) argument $z$; and $C(z)$ denote the set of contexts occurring at some point in token argument $z$. In what follows, let the predicate of token arguments, STAB$(x)$ express that the extensions of the (non-null) set of expressions $E(x)$ are invariant across the (non-singleton) set of contexts $C(x)$. Let GOOD$(x)$ say that the token argument $x$ is in logical good standing. EI and EII express extreme inscrutability and extreme indexical inscrutability respectively. Let $D$ be the operator 'determinately'

The master argument against extreme inscrutability can be formulated as follows, where $A$ is some intuitively 'logically good' token argument:

| | | | |
|---|---|---|---|
| 1. | $D(\text{GOOD}(A))$ | (premise) | (1) |
| 2. | $D(\text{GOOD}(A) \supset \text{STAB}(A))$ | (premise) | (2) |
| 3. | $D(\text{GOOD}(A)) \supset D(\text{STAB}(A))$ | (D-distribution) | (2) |
| 4. | $D(\text{STAB}(E(A), C(A)))$ | Modus Ponens, 1,3 | (1,2) |
| 5. | EI $\supset$ EII | (premise) | (5) |
| 6. | EII $\supset \neg D(\text{STAB}(E(A), C(A)))$ | (premise) | (6) |
| 7. | $\neg$EII | Modus Tollens, 4,6 | (1,2,6) |
| 8. | $\neg$EI | Modus Tollens, 5,7 | (1,2,5,6) |

Thus, we have a case against inscrutability, on the basis of just four premises. The first it seems to me, is undeniable: some token arguments we offer are (determinately) in *logical* good standing.[24]

The second premise is yet to be argued for. It is the claim that, for an argument to be logically good, there must be no change of extension through the contexts at play in the argument (and that this is determinately the case). The burden of the second half of the paper is make the case for this.[25]

The third premise ((5), above) is something for which an argument has already been given, and which is further elaborated in the appendix. This is that *if* extreme inscrutability is the case, then extreme indexical inscrutability is the case. I base this not on the grounds that there is some *entailment* between the two claims (and note that the material conditional suffices for my purposes). Rather, it is supported on the grounds that the *argument* that is put forward to support extreme inscrutability generalizes to establish extreme indexical inscrutability.

The fourth premise ((6), above) is that indexical inscrutability means that there is no fact of the matter about whether or not the intensions of expressions involved in the argument remain constant throughout the contexts involved in the argument. This follows immediately from the definition of indexical inscrutability.

The argument is clearly valid, involving only Modus Ponens, Modus Tollens, and the principle of D-distribution (an instance of the modal logical principle '$K$' for the 'determinately' operator). All that remains, to close the case against extreme inscrutability, is to establish premise (2).

As with the arguments previously sketched, we consider whether the argument will impact on *moderate* inscrutability just as much as *extreme* inscrutability. For definiteness, suppose one were to reject extreme inscrutability, but hold, with Quine, that 'rabbit' is indeterminate between a small *range* of rabbit-candidates (rabbit slices, undetached rabbit parts, etc). Let *moderate indexical inscrutability* in this setting be the view that it's indeterminate from one context to the next whether or not the reference of 'rabbit' changes from of these candidates to another, or rather stays the same. Can we replace EXTREME INDEXICAL INSCRUTABILITY and EXTREME INSCRUTABILITY throughout with MODERATE INDEXICAL INSCRUTABILITY and MODERATE INSCRUTABILITY, and thus get an argument against even moderate inscrutability of reference? We then might get an argument for total determinacy of reference: a claim that would be of interest far beyond the present debate.

I take it that premises 1,2, and 6 are just as good on the new argument as on the old (in particular, moderate indexical inscrutability is just as good as extreme indexical inscrutability at undermining determinate stability of the expressions in an argument). The real question is: can we defend the material conditional from straight moderate inscrutability to moderate indexical inscrutability?

I see no general reason to think that moderate inscrutability *in the absence of extreme inscrutability* will generalize in this way. For if we have moderate inscrutability *without* extreme inscrutability, then the facts that fix meaning must be more extensive than the bare interpretationist allows. We may again take as representative of such a theory Lewis's insistence that eligibility, as well as fit with sentential data, needs to be optimized by the meaning-giving interpretation of natural language. We are allowed to assume, for present purposes, that rabbit-stages, rabbit-worms and the rest, each lead to equally eligible overall interpretations of the language.

Now, all these interpretations will make 'Rabbit' no more indexical than the English expression 'Rabbit'. The Kaplanian *character* will be constant: assigning to the predicate the same content no matter what the context is. If Lewis's eligibility constraint only requires eligible properties to be assigned *in each context*, then moderate indexical inscrutability may follow. However, I contend that Lewis's theory should directly assess the eligibility of *characters*.[26] It is then plausible that the function from contexts to intensions that assigns to 'Mopsy' a certain rabbit-stage in one context, a certain four-dimensional entity in the next, and a certain

undetached rabbit-part in the third, will thereby be objectively *less eligible* than a function that assigns to 'Mopsy' one and the same rabbit-stage at all contexts. On at least one interpretationist account that denies extreme inscrutability but leaves room for moderate inscrutability, moderate indexical inscrutability will not follow. The moral is more general: those constraints that rule out the 'extreme' deviant interpretations may well rule out the kind of radically inconstant interpretations needed for moderate indexical inscrutability. In the nature of the case, it must be determined on a case-by-case basis, whether or not a given account of moderate inscrutability is or is not committed to moderate indexical inscrutability, and thus susceptible to the same criticisms here offered against extreme inscrutability. But unlike the extreme case there is no *general* argument in the offing against moderate inscrutability.)

# II

The master argument against extreme inscrutability of reference will go through if we can defend the last remaining premise: the principle that the language involved in logically good token arguments must be *stable* throughout the contexts involved in the argument. For any token argument $A$, we need to establish the determinate truth of the conditional:

$$\textsc{good}(A) \supset \textsc{stab}(E(A), C(A))$$

My case for this is that we need to endorse the conditional if we are to address a *general* puzzle about the relevance of logical consequence to token inferences carried out in natural language.

The starting point for the puzzle is the relationship between a certain kind of virtuous epistemic practice, and the classification of argument *types* as logically valid or invalid.[27] The virtuous epistemic practice is that of competent deductive inference.

Competent deductive inference is one of the central ways in which we extend our knowledge. This is reflected in the vital role it plays in central epistemological principles such as closure:

> Necessarily, if $S$ knows $p$, competently deduces $q$, and thereby comes to believe $q$, while retaining knowledge of $p$ throughout, then $S$ knows $q$. (Hawthorne, 2004, p.34)[28]

If competent deduction is an epistemically virtuous way of reasoning, why does it have this status? Well, it is a reliable form of reasoning, unlikely to lead us from truth to falsity. But all sorts of other forms of reasoning are reliable which seem not to have the distinctive standing of competent deduction (induction is a case in point).

What seems to be distinctive about competent deduction is that it is *logically good*: the conclusion *follows from* the premises. It is the fact that competent deductions are *tokens of a valid type* that separates them from induction or to other generally reliable heuristics. (Of course, there are plenty of other constraints on *competent* deduction beyond simply being logically good. For example, perhaps a competent deducer needs to be *aware* of the logically good status of the argument he is giving. I take no stance here on the nature and extent of these further conditions that make logically good arguments into competent deductions).

It is one thing to point to the feature that separates competent deductions from other forms of virtuous reasoning, however, and quite another to *explain why that feature is epistemically relevant*. Thus, there is in-principle a gap between the descriptive claim that a competent deduction is of a logically valid *type* and the normative claim that it is thereby epistemically *good*.

At first glance, it seems no mystery that type-validity is a good-making feature of token arguments. For validity, of course, guarantees that whenever the premises of the argument are true, the conclusion is true. Thus, it seems that type-validity gives a *logical guarantee* of reliability; and that seems at least a *pro tanto* good.

Our puzzle strikes at exactly this point. *Pace* the above, when we consider token arguments in natural language, we shall see that there is no general guarantee of reliability. Absent further explanation, this makes a mystery of why competent deductions should be so epistemically special: for the explanation of the epistemic relevance of their only distinctive feature has lapsed.

In what follows, I begin by setting out some initial difficulties in extending the notion of validity to natural language. There is a standard Kaplanian resolution of these difficulties, but as I shall explain the resolution generates exactly the puzzle just flagged: it makes the epistemological relevance of logical consequence obscure. We can resolve this difficulty, I

argue, but only if we say that *more* is needed for an argument to be in logical good standing than that it simply instantiate a valid type: we need to also assume the kind of *stability* of language through the contexts of the argument discussed earlier. Thus, resolving our puzzle requires us to endorse the above conditional: the final premise in the master argument against extreme inscrutability.

# 6   Validity for indexical languages

Tarski famously showed how to make *logical consequence* formally tractable. He did this by analyzing it via what we now call *semantic* or *model theoretic* validity. At a first approximation, something is semantically valid if *no matter what the subject matter of the premisses and conclusion might be*, the inference is truth preserving. For example, the argument 'Billy is running; therefore something is running' will count as valid in the intended sense, since no matter what "Billy" refers to, or what "is running" means, the truth of the premiss will secure the truth of the conclusion.

This basic idea is cashed out more formally as follows: an argument $p_1, \ldots, p_i, \ldots,$ *therefore q* is semantically valid if, on every *admissible interpretation* of the language, either one of the $p_i$ is false or the conclusion is true.[29] Grades of semantic validity will emerge, depending on what we count as an admissible interpretation. If the sole constraint on admissibility is that the interpretation of logical connectives be held constant, as in the "Billy" example above, then we get a notion of formal or logical consequence. More severe constraints on admissibility lead to different consequence relations. The semantic treatment of consequence does not enforce any particular choice of admissibility constraints: what choice is appropriate can be left for independent argument to determine.[30]

So long as we can classify sentences as true or false *simpliciter*, the above definition of semantic validity makes straightforward sense. It is a familiar fact about natural languages that (unambiguous) sentence types can change their truth-value on different occasions of use. We cannot assign a truth-value directly to the sentence "I am George Bush": for it is true when uttered by Bush, and false when uttered by me. Similarly for temporally indexical sentences "it is now 5 o'clock": true as uttered at 5pm, false as uttered an hour later.

Accordingly, even the most obviously 'valid' inference-patterns will on occasion fail to preserve truth, in an indexical language. The rule of reiteration *from S, infer S* is obviously valid for non-indexical languages; but if we instantiate the pattern with "it is now 5 o'clock", and allowing the two tokens of *S* to take place an hour apart, it will clearly fail to preserve truth.

This, then, is the puzzle with which we begin. When we consider indexical languages, even the very simplest valid argument-types (reiteration, conjunction elimination, etc) have tokens which have true premises and a false conclusion.

This generates two challenges. Semantic validity was defined in terms of *truth preservation*, and we have seen that valid arguments sometimes *fail* to be truth-preserving. So we need to show how to characterize a notion of semantic validity for argument *types* for an indexical language, which such examples do not immediately render trivial. The second challenge goes in the other direction: given a characterization of semantic validity which allows basic argument types such as reiteration and conjunction elimination to count as valid, we will have to admit that some tokens of valid argument types do not always preserve truth. So we face the burden of explaining why validity is relevant to inferential practice at all.

## Kaplanian validity

Kaplan's paper "Demonstratives"[31] tackles the first challenge. The first step towards systematizing validity for indexical languages is to characterize sentences, not in the first instance as plain *true* but rather as *true in the context c*.[32] For example, "it is now 5 o'clock" is *true in the context $c_1$* (where the time of $c_1$ is 5pm) but *false in the context $c_2$* (where the time of $c_2$ is 6pm). We can describe an utterance of the type *S* as plainly true if the sentence is *true in the context c*, where *c* corresponds to the setting in which this token of *S* is uttered.[33]

The basic characterization of semantic validity does not have direct application to languages containing a parameterized truth predicate "true-in-c". At the time at which Kaplan's work was written, there were extant techniques for extending the semantic characterization to parameterized truth-predicates. If we are interested in the *truth-conditions* of sentences—their truth-values in different possible situations—then 'possible world semantics' involves a relativized notion of truth. According to possible world semantics, a sentence such as 'Billy is sitting' is not in the first instance true or false *simpliciter*, it is true or false *at possible world w*. (Again, we can derivatively assign truth *values* to token sentences—as uttered in a concrete situation, *S* will be plain *true* if it is true *at the actual world*.)[34]

As in the case of indexical languages, the resulting parameterized truth-predicate blocks the straightforward transfer of the characterization of validity in terms of unrelativized notion of *truth*-preservation. In the intensional case, the answer was familiar by the time that Kaplan was writing "Demonstratives": we characterize validity in terms of *truth-at-w* preservation, at all points *w*, under any admissible interpretation.[35] That is, we first say what it is for an argument to be *true-preserving at all points*: for every point *w*, either one of the premises is false-at-*w*, or the conclusion is true-at-*w*. As a second step, we generalize over interpretations in the way characteristic of semantic definitions of validity.[36]

Kaplan's "Demonstratives" suggested that we apply the same trick to indexical languages. Again, we find a parameterized truth predicate, truth-at-context-*c*.[37] To define validity, we must first 'generalize away' this parameter. First, we say what it is for an argument to be *truth preserving at all contexts*: every context *c* such that the premises are true-at-*c* must be such that the conclusion is true-at-*c*. Then, as a second step, we generalize over interpretations in the way characteristic of semantic definitions of validity. The result is that an argument is valid if it is truth preserving at all contexts, under any admissible interpretation.[38]

Kaplan's treatment gives us a definition of semantic validity for indexical languages. Significantly, it shows how such a notion can be non-trivial. For an argument to be valid, it is required only that it be truth-preserving at each particular context. It does not require that it be truth preserving when the premises and conclusion occur in different contexts. Hence, reiteration can be a valid rule, and we are allowed to dismiss as irrelevant 'counter-examples' to its validity involving change of contexts.

## Problems

I will assume that the Kaplanian treatment of indexical languages is the correct way to answer the first challenge: to characterize a non-trivial notion of *valid argument type* for indexical languages. Defining validity in this Kaplanian way leads to an immediate puzzle about the relevance of the validity of an argument type, so-characterized, to the practice of inferring one thing from another.

In the basic case of a non-indexical, extensional language there is no gap between validity of patterns of inference (inference-types) and the good standing of concrete exemplifications of that type. Since validity guarantees that the inference in question preserves truth *simpliciter*, the

information that inference-pattern $I$ is of a valid type assures us that in any token of type $I$, the truth of all the premises ensures the truth of the conclusion.

We find a similar situation for the intensional case. If $I$ is type-valid, then we know that the inference is truth preserving in every possible situation. Obviously, any token of that type with which we would be concerned, would take place *within* a single possible world, so as a special case we have it that such tokens of that type will preserve truth.

In the case of indexical languages matters are not so straightforward. *If* the whole of a token inference occurred in some particular context $c$, then since a valid inference-type is truth-preserving in every context, it is truth preserving in *the one in question*. The situation would be exactly analogous to the intensional case above. However, token inferences—strings of utterances or inscriptions[39]— *prima facie* involve a series of distinct contexts. It takes time to utter or inscribe successively the premises and conclusion of a token inference. If the premises and conclusions occur in distinct contexts, there is no guarantee that the token inference will preserve truth, even if it is of a valid type.

Again, we can illustrate this with the reiteration rule. Consider the inference from "Jim walks" (uttered at $c_1$) to "Jim walks (uttered at $c_2$). The inference-type here instantiated is clearly valid, on Kaplan's characterization. At any single context, under any re-interpretation, it will preserve truth. However, we can find an interpretation of the vocabulary that will make the premiss true-at-$c_1$ and the conclusion false-at-$c_2$. (Example: Consider the re-interpretation according to which "walks" is an indexical predicate designating *walking* at $c_1$ and *skydiving* at $c_2$.) We have *an* interpretation which makes this combination of inference-type and change of context non-truth preserving. Absent further information, one can conclude *nothing* about the truth-preservingness of the actual inference one makes, from the fact that the inference-type tokened is semantically valid in Kaplan's sense.

# 7   The modest proposal

The threat is radical. The validity or otherwise of inference-types threatens to become irrelevant for the practical purposes of reasoning. *Prima facie*, all reasoning involves changes of context— and any such change would seem to render the Kaplanian treatment of validity irrelevant. Our challenge is to explain the *relevance* of information about validity of inference-types to the rightness or wrongness of specific inferential practice.

One reaction is to conclude that we should be interested in some non-semantic notion of valid inference-type. I think that this would be overhasty. There is a modest solution at hand, that shows how the semantic validity of inference-patterns can still have practical significance.

The modest idea is the following. Call a token inference *good* if (1) it instantiates a valid pattern in Kaplan's sense (2) no relevant change of context occurs during the course of the inference. The contention is that the information that a token inference's good standing, in the above sense, guarantees that it preserves truth, and hence will justify inferential practice. Since part of the characterization of good standing is the validity of the argument type, we explain why information about validity is relevant to correct inferential practice.

Note that the characterization of good standing allows *irrelevant* changes of context to take place over the course of an inference. Suppose we make an arithmetical inference. Then, intuitively, no change of context will be relevant to this argument token. Whatever change of context occurs, the argument will be good.

### Stability, Invariance and Constancy

Which token inferences are classified as good will depend on how we cash out the notion of 'relevant change of context'. The basic idea here is that a change of context will only be relevant if it changes the intension of some expression occurring somewhere in the inference. Note that though condition (1) is, by its nature, appreciable independently of what the particular meanings of the words are, appreciation that (2) obtains will essentially depend on what indexical character the words have.

Let $\tilde{C}$ be the set of all contexts involved at some stage of a token inference $i$. If $i$ is of a valid type, then so long as no relevant change of context occurs, we can be sure that we will not be led from true premises to a false conclusion. For we know, by condition (1), that within a single context this can never happen. Now consider the correct interpretation of the premisses; this will be some $c \in \tilde{C}$. Hence we know that the conclusion is true as uttered in $c$. However, any legitimate reading of the conclusion will be assessed relative to some context $c' \in \tilde{C}$. By condition (2), we know that $c$ and $c'$ assign the same content to the expressions that feature in the inference (for otherwise, a relevant change of context would have occurred). By compositionality, an assignment of intensions to the primitive parts of a sentence will fix the intension of the sentence as a whole. Hence we know that the conclusion has the same intension when assessed relative to $c$ and $c'$. Sentences with the same intension are true in exactly the same situations. Since the conclusion is true in $c$, it is true in $c'$ also.[40]

The basic idea that *a change of context will only be relevant if it changes the content of some expression occurring somewhere in the inference* needs more exact formulation. I provide two alternative ways of precisifying this notion, formulated for the general case of an intensional, indexical language.

**INVARIANCE** Call a set of contexts $\tilde{C}$ *invariant* with respect to a set of expressions $E$, when each expression in $E$ is assigned the same intension at every context in $C$.

**CONSTANCY** Call a set of contexts $\tilde{C}$ *constant* with respect to a set of expressions $E$, if any aspect of context that is used to fix the intension of a term in $E$ is common to every context in $\tilde{C}$.

Compare these to the following:

**STABILITY** A set of contexts $\tilde{C}$ is *stable* with respect to a set of expressions $E$, when each expression in $E$ is assigned the same extension at every context in $C$.

Notice that constancy entails invariance which entails stability (sameness of aspects of context relevant to determining intension guarantees sameness of intension, which guarantees sameness of extension). Of course, the stability, in this sense, of the expressions and contexts in play in token argument $A$ is just what is expressed by STAB($A$).

Either invariance or constancy could be used as an explication of condition (2), but either way, a good argument has to be such that the relevant set of contexts is *stable* with respect to the expressions involved, precisely in the sense relevant to the truth of the conditional needed to complete the master argument against extreme inscrutability.

### Review

The central idea here is to limit the *range of relevant changes of context* that we need to worry about when evaluating a certain inference. If we can legitimately presuppose that context does

not change in any of those ways, then the fact that the inference is a token of a valid type will guarantee that truth will be preserved through the inference. An attractive feature is that the only knowledge that we need, in principle, in order to work out what kind of context changes are relevant, will be basic linguistic knowledge about the meaning of the expressions used in the inference.

# 8   Objections and responses

Objection 1.

Constancy is too demanding to play the role envisaged above. For context is constantly changing: utterances of premises and conclusions are never totally simultaneous, for example. No inference involving any temporal indexical will ever correspond to a constant set of contexts.

Reply 1.

There are several responses available, getting progressively less concessive.

(a) The most concessive response is to grant that no inference involving (for example) temporal indexicals will *ever* be in good standing, as characterized above. For recall that the aim was never to extend greatly the range of arguments we could legitimately declare valid. Rather, its goal is to save what we already think we have: to explain how the notion of semantic validity can have practical significance in application to uncontroversially valid patterns of inferences carried out in apparently unobjectionable settings.

For example, correct inference probably plays a central role in the epistemology of mathematics. Mathematical English is, however, just an indexical-free fragment of a broader indexical language ("$2 + 2 = 4$ and I am here" is intuitively well formed). As part of an indexical language, mathematical English, though it contains no indexicals, is subject to re-interpretations wherein its terms are treated as indexicals. Token mathematical inferences are therefore under threat. However, constancy and invariance, are trivially secured for tenseless, indexical-free fragments of language, such as the language of pure mathematics: this captures the sense in which consideration of indexical re-interpretations of the words seems so *irrelevant* to the good standing of mathematical inferences.

Beyond this, we are used to the idea that we might need to substantially reformulate arguments in order for them to meet the strictest standards of validity—and reformulation in tenseless language might be one of these cases.

(b) A less concessive response is to preserve the potential good standing of token inferences involving temporal indexicals by characterizing the notion of good standing in terms of invariance rather than constancy. The temporal indexical used in inferring "Today is Tuesday" (at 13.01 hours) from "Today is Tuesday" (at 13.00 hours) does not render the relevant set of contexts variable, though it arguably renders them inconstant.

(c) Less concessively still, one can question whether the context change involved in the above inference does render the set of contexts inconstant. On one view, contexts comprise a string of factors: the time, place, agent, world, etc. at which the utterance takes place. On this picture, the character of the indexical "Today" would be something like

"the day containing the temporal component of context $c$". There is then a relevant difference in context during the above inference; and hence the set of contexts associated with the inference would be inconstant. The current objection tacitly presupposes this kind of setup.

An alternative is to represent contexts as centred worlds, and appeal to auxiliary functions that pick out aspects of those worlds. For example, the interpretation of 'today' would be given by ' $\text{DAY}(c)$'. The 'aspect' of the context that is required to be invariant in the case of the "Today" inference, would be $\text{DAY}(x)$ for $x \in \tilde{C}$. This kind of setup is attractive, if only because it seems hopeless to list once-and-for-all all the possible aspects of context that could be relevant to determining the reference of an indexical expression.[41]

If this is the case, then one can analyze goodness in terms of constancy, while still classing the 'Today' inference as good.

Objection 2.
    The account fails to capture the interesting logical or conceptual connections between indexical utterances uttered in different contexts. From 'it is the case that $F$' we can legitimately infer, at a later point, 'it was the case that $F$'. This inference is not valid in Kaplan's sense, and so the token will not be in good standing, in the current sense.

Reply 2.
    It is not my intention to say anything novel about such questions at this stage. As outlined in the response (a), above, the radical threat in view is that the Kaplanian notion of validity turns out to be *irrelevant* to actual inferential practice. Indexicality threatens the application of the concept of validity to *all* token inferences. Faced with this threat, the first job is to secure the good standing of the boring, run of the mill token inferences that we all take for granted.

This is not to deny the interest of more ambitious projects that try to explain the distinctive character of inferences involving indexicals that do not obviously instantiate a valid inference type—and indeed, ones whose 'validity', if such it is, *depends* on a change of context. I contend (but will not defend here) that an account of such 'dynamic inferences' can be made out that dovetails with the present account: we can outline a chain of good inferences in the sense of this paper, that take us from "Here, it is raining" to "There, it is raining" as one drives out of and away from a thunderstorm while keeping track of its location; We can exploit similar stories to account for good and bad cases of basing one's utterance "it was raining" on an earlier utterance of "it is raining".[42]

The account given here is not committed to the success of these more ambitious projects. Nevertheless, the availability of an extension of the approach to these more controversial cases guards against accusations that we are attempting to theorize about good inference using the wrong tools (e.g. materials from the theory of reference, rather than the theory of sense).

Objection 3.
    The account fails to cover some obviously valid patterns of inference. For example, consider the inference from 'it is now 13.46.24 precisely' to 'everything is self identical'. The argument is clearly a good one, in virtue of the consequent being a logical truth, but since the temporal component involved in the antecedent will alter through the course of the inference, it will not be 'good' in the sense sketched above.

Reply 3.

To begin with, the now familiar point about the aim of the project being to save the applicability of uncontroversially valid inference types, rather than to give a sense in which more controversial types are valid, is again relevant. Let us set this aside for the moment, however.

In the particular case cited, we might be able to modify the definitions of invariance or constancy to get the result the objector wants. We could, for example, demand constancy only of aspects of context that are appealed to more than once in the course of an inference. However, new examples would be constructed making the same point: consider 'it is now 13.46.24 precisely' to 'everything is self identical or it is now 13.46.24 precisely'. Clearly, more and more fine grained notions of constancy could be developed, seeking to capture the idea that there should be no change in aspects of context which are *essential to the validity of the inference*. This looks like a project that will have diminishing returns.

Better to bite some bullets. Let us accept *pro tem* that the inference is not good, in the relevant sense. What bad results follow? Notice that we can characterize a derivative sense in which the inference is well-made. For example, the inference from the null set of premisses to 'everything is self identical' is valid. If we know this, and also know that (quite generally) a token argument will never take us from truth to falsity if it is a *weakening* of one that is in good standing, then we know enough to explain how we appreciate the truth-preservingness of the original argument. I anticipate telling similar stories, on a case-by-case basis, for other examples adduced.

Objection 4.

You haven't given what you promised! The notion of semantic validity still has no application to token inferences.

Reply 4.

I have chosen to introduce the neutral notion of the *logical good standing* of a token inference, rather than to describe this status as *validity*. This is because I do not wish to be involved in disputes that seem to me to obscure the central point at issue.

One might think that in cases where a token inference is performed where there is change in an aspect of context that affects the reference of one of the terms involved, it would be a category mistake to apply the notion of 'validity' to the inferences in question—evaluation of an argument as valid or invalid occurs only *after* we have ensured against 'ambiguity'-like phenomena in the expressions used. Alternatively, one might want to characterize an inference from "it is now 5pm" (uttered at 5) to "it is now 5pm" (uttered an hour later) as *invalid*.

I take no stance on the issue, which strikes me as essentially terminological. Depending on which approach one favours, one can construct the relevant notions out of the materials provided–in particular, the neutral notion of a good inference. The key feature is to explain how appreciation of the semantic validity of arguments could be practically relevant; and this has been explained, if it is granted that the good standing itself is practically relevant.

Objection 5.

We need not think of token inferences as involving change of context at all. The modest proposal is therefore unnecessary: semantic validity will suffice.

Reply 5.

This is a view suggested by Williamson (1997). After describing the Kaplanian treatment of validity for indexical languages, Williamson notes the challenge:

> There is a problem. On this account, the validity of an argument tells us nothing about truth-preservation when the premises and conclusion are instantiated in different contexts. However, inference takes time: we judge the premises and conclusion successively. Thus the account of validity looks irrelevant to the inferences we actually make. (Williamson, 1997, p.652)

Williamson holds that the problem only arises if we focus on the relations between occurrent *judgements*. Plausibly, it takes time to move from judging (or supposing) that premises hold, to judging that the conclusion follows. However, Williamson contends that we should properly be focusing on *belief states*, which can be retained over time. These simultaneous belief states can be considered as instantiating, or failing to instantiate, a valid pattern of inference; and no concern about the involvement of different contexts emerge.

I think two points show Williamson's approach does not get the heart of the issue. Let us suppose *pro tem* that his suggestion handles *cross-temporal* context change. Nevertheless, there are other aspects of context that may still vary between premises and conclusion of an inference. We can bring this out with a simple, though admittedly artificial, indexical expression. Let me introduce the indexical 'blig', which will refer to the primary connective of any sentence or sentence-like structure in which it figures, and to the Taj Mahal otherwise. 'Blig is the Taj Mahal' is true; but 'Blig is the Taj Mahal and Blig is the Taj Mahal' is false.

The point is *the placement of the indexical itself* is potentially an aspect of the context with respect to which the indexical is assigned a referent. Though the example is artificial, it would be most uncomfortable to exclude expressions with this feature *a priori*. Indeed, more standard philosophical examples might well be analyzed as involving expressions of this type: e.g. self-referential sentences or thoughts such as "this very sentence is short", which presumably rely on aspects of context such as *the demonstrated sentence*. A related point is that context can change *within* a sentence, as Lewis (1979, p.241) emphasizes. If the above proposals are accepted, we look in danger of developing a notion of context and inference that is unable to discharge its other theoretical roles. We need to accommodate such context change, not eliminate it.

Further, the account stands in need of an extension to handle ascriptions of validity to sequences of occurrent judgements, written and spoken inferences, and the like. We *do* evaluate such sequences of events as valid or invalid, and we need a reconstruction of this practice. Since the issue of cross-context inferences has not been taken head-on by the Williamson proposal, we have no ready-made proposal for such cases. Presumably, such ascriptions of validity must be handled derivatively from the primary notion of a valid inference involving belief states—but we have been given little indication how this is to be effected.

Objection 6.

There is an obvious response that the current discussion misses. We should distinguish the *context in which a judgement (or utterance) is made* from the *context with respect to which a judgement (or utterance) should be assessed*. Accordingly, we should think of a token inference as assessed at for a single context—though it is indeterminate which

one this should be. Since semantic validity guarantees truth-preservation in all contexts, it guarantees truth-preservation no matter which of these contexts we take to be the actual one.

Reply 6.

This strategy is canvassed by both Campbell (1997) and Williamson (op cit). However, it is easy to find examples where it delivers inappropriate results. Consider an argument of the following form:

This very second is an even number of seconds after *t*

This very second is an even number of seconds after *t*

Now, if some time has elapsed between the judgement that the premiss holds and the drawing of the conclusion, it is clear that the 'argument' by reiteration should not be taken to be in good standing. However, every choice of context will declare it valid, since a choice of context will determinate a time which will be the referent of 'now' in each case. Assessing the token argument with respect to a single context delivers exactly the wrong result: this is an instantiation of a valid argument type that we want to come out invalid.

The point becomes more dramatic if we use the indexical 'blig', introduced above. Consider the following argument:

Blig is the Taj Mahal

so:

Blig is the Taj Mahal and Blig is the Taj Mahal

The premise is unambiguously true (in a sentence with no main connective, 'blig' will designate the Taj Mahal); the conclusion is unambiguously false (in a conjunctive sentence, 'blig' will designate Conjunction). With respect to any *single* choice of context, 'blig' will have to designate a single object; and hence the argument will come out as truth preserving. Again, the argument-type is truth preserving, but no argument token should be described as in good standing.

The point is that the separate elements of the token argument—the token premiss and conclusion—call for distinct contexts. We cannot force them into a single context without clear distortion of the intended use of the expressions. There just is no admissible single context for the token inference.

We were looking for conditions that articulate what it is for a token argument to be in *logical* good standing. Being a token of a valid type is not sufficient. I have argued that the modest proposal resolves our puzzle here. It does not lead to unacceptable consequences (*pace* the objections above); and moreover, the rival suggestions proposed by Williamson and Campbell fail.

The modest proposal neatly resolves a puzzle about the epistemic relevance of type-validity. But in the present context it has a wider significance. For, by requiring that expressions in token arguments be *constant* or *invariant* through the contexts in play, it requires *a fortiori* that they be stable, in the sense that their extensions remain the same throughout those contexts. Determinately, if *A* is a logically good token argument, then it is stable.

# III

Extreme inscrutability is counterintuitive. But it is a consequence of an otherwise attractive theory of linguistic content: an elegant form of interpretationism. That motivates the project of weighing the benefits and costs of this metasemantic package.

Flat footed appeals to intuitions are dialectically weak, teach us little, and such intuitions may in any case be 'explained away'. But it turns out to be surprisingly difficult to pin down precise theoretical costs that the inscrutabilist faces. I argued in the first half of the paper that there are *prima facie* lines of resistance to the central lines of argument one might try against inscrutabilism.

The master argument of this paper, however, gains traction against the inscrutabilist. It shows that there is a tension between extreme inscrutability (adopting on the basis of permutation-style arguments) and the epistemology of inference. There are two key premises to the master argument. The first is that extreme inscrutability generalizes to extreme indexical inscrutability. This is *not* immediate (in particular, there is no logical entailment from the former to the latter). But we can argue for the truth of the needed material conditional. The argument was sketched in the text, and is developed in formal detail in the appendix. (Importantly, there is no general reason to think the analogous conditional holds when it is *moderate* rather than extreme inscrutability that is at issue).

The other key premise is that, determinately, token arguments which are 'logically good' involve no change of extension across the contexts they cover. The second half of the paper argues that, in order to resolve a puzzle over the epistemic relevance of (type) validity, we need to analyze 'logically good' in a way that secures this conditional.

The master argument shows that from the determinate logical good standing of any token argument, one can deduce the falsity of extreme inscrutability. Accordingly, the price of inscrutability is that no token argument is in determinate good standing. One might be willing to pay much to secure a workable reductive theory of linguistic representation: but the price that (extreme) inscrutability extracts is too high.

# Appendices

In the following appendices I provide the formal overgeneration results referred to in the main text. I show that these results hold in the kind of setting used by many linguists to provide a semantic theory for natural language: the generalized semantics of Lewis (1970). In appendix A I develop the tradition permutation arguments in this highly general setting, and respond to recent contention by Vann McGee (2005a) that permutation arguments for inscrutability *break down* when we attempt to generalize them to rich languages. In appendix B I show how to prove the extreme indexical inscrutability result at the heart of the criticisms of inscrutabilism developed in this paper.

## A   Permutation arguments preserving truth-conditions

In some interpretationist settings, one's data consist of a pairing of sentences with coarse-grained propositions or intensions (cf. the interpretationism described in Lewis (1975)). The target is to preserve, not only the truth-values of certain sentences, but the *truth-conditions* of sentences. In the current setting we shall think of truth-conditions as given by sets of possible worlds, or characteristic functions of such sets: this tells us *the worlds in which the sentence is true*. Of course, if we can show that permuted reference schemes assign the same *truth conditions* to sentences, then it will follow immediately that they assign the same *truth value at the actual world* to sentences.[43]

We will show how arbitrary permutations of the extension of singular terms can be embedded within an overall interpretation which leaves the truth-conditions assigned to sentences invariant. In fact, the setting is more general. The semantics will assign to sentences functions from indices to truth-values, where the indices specify not only a possible world, but also a time, place, and other factors. As explained at §A, if we can show that the permuted variants of an interpretation leave the semantic value of sentences invariant, we shall have what we need.

We first prove the result for the general semantic setting of Lewis (1970) (what Cresswell (1973) calls a 'pure categorial language'). We first set out the framework itself, and then state and prove a permutation theorem for the setting. We shall then describe how the framework is altered to take into account indexicality, and extend the permutation argument to that 'double-indexed' setting.

**The framework (syntax): a pure categorial language**

The framework in which we operate initially is a *general semantics*, in the sense of Lewis (1970). The atomic expressions of the language will be given by a *lexicon*, and there will be phrase-structural rules saying how admissible compounds can be built up from elements of the lexicon. The lexicon and the phrase-structural rules give the analogue of a definition of well-formedness within a formal language.

The framework that Lewis recommends is drawn from Ajdukiewicz's (1935) categorial grammar, and is what Cresswell (1973) calls a 'pure categorial language' (essentially, this is a form of type theory, without any 'syncategorematic' terms such as λ-operators). In the lexicon, atomic expressions will be associated with *categories*, and then the well-formedness of compounds will be determined by whether the categories of the components dovetail. For example, the category of the lexical item "Susan" might be that of 'names' ($N$), and the category of the lexical item "runs" might be 'intransitive verb' ($S/N$)—the notation reflecting the fact they when combined with a category $N$ expression it produces a category $S$ expression (a sentence).

The categories themselves are characterized recursively. $N$ and $S$ are 'basic categories'[44] Given categories $c, c_1, \ldots, c_n$, we have a *derived* category $c/c_1, \ldots, c_n$. In the above, 'runs' is in the derived category $S/N$. Writing "+" for concatenation, the general rule of well-formedness is that for any expressions $e, e_1, \ldots e_n$, we have $e + e_1 + \ldots + e_n$ well-formed iff $e$ is of a derived category $C/C_1, \ldots, C_n$ and each $e_i$ is of category $C_i$. "runs(Susan)" is well-formed because it is of the form $(S/N) + N$, and so fits our template.[45]

Notice that the complexity of the category of an expression and its syntactic complexity cross-cut. Basic lexical items—the atomic expressions—will typically have quite complex categories: for example, the atomic adverb 'fast' will have category $(S/N)/(S/N)$ (it yields a verb phrase when a verb phrase is input). Conversely, expressions of the basic category $S$ (sentences) will be syntactically complex.

The result of such concatenation won't look much like English. A very basic example is that 'Jill loves Jane' would be represented by the word-order '(loves(Jane))Jill' within the system just developed. Additional work is required to explain how such analyses relate to natural language. What we need here is some kind of mapping from our 'disambiguated' formal language to ordinary sentences of English. This 'ambiguating relation' may take one of a number of forms, depending on what other resources from linguistics one wishes to bring to bear. Lewis (1970, p.204) appeals to a "transformational component" of the grammar; Dowty (1979) suggests various ways to integrate the setting with generative semantics.[46]

### The framework (semantics): single-indexed general semantics

What we have so far seen is a treatment of the syntax of a formal language. We now describe the Lewisian "general semantics" which will associate each well-formed expression with a semantic value. The kind of semantic value for a given expression will be what Lewis calls an "appropriate intension"—which intensions are appropriate being determined recursively by the category of the given expression. The base cases are handled individually. For sentences (expressions of category $S$) an appropriate intension will be a function from indices to truth-values. For expressions of category $N$, it will be a function from indices to objects.[47] Appropriate intensions of derived categories are then functions between intensions appropriate to the categories from which they are derived. For example, expressions in category $S/N$ will be assigned functions from intensions appropriate to category $N$ expressions to intensions appropriate to category $S$.

We can capture this by characterizing intensions in terms of their *type*. Let indices be of type $i$, objects type $o$ and truth-values type $t$.[48] Let the type $\langle a, b \rangle$ include all functions from elements of type $b$ to elements of type $a$. We can then formulate the above by saying that an appropriate intension for category $S$ is a function of type $\langle t, i \rangle$; an appropriate intension for the category $N$ is of type $\langle o, i \rangle$; and in general the appropriate intension for category $c/d$, where $c$ is of type $\alpha$ and $d$ is of type $\beta$, will be of type $\langle \alpha, \beta \rangle$.[49][50]

A *lexical interpretation* $||$ will be an assignment of semantic values to basic parts of the language—the elements of the lexicon (e.g. names, adjectives, verbs, adverbs etc). We now have the idea of a lexical interpretation function for a categorial language, and a characterization of the *type* of semantic value appropriate to complex expressions. We need to 'extend' the lexical interpretation function to give an overall assignment of semantic values to arbitrary well-formed expressions in the language.

We let a model $m$ for the language include a domain $U$, a set of indices $I$, and a lexical interpretation function $||$. Then one could let a model induce a valuation $v_m$ for arbitrary well formed expressions:

1. For $e$ in the lexicon, $v_m(e) = |e|$

2. For $e' = e(e_1, \ldots e_n)$, $v_m(e') = v_m(e)[v_m(e_1), \ldots, v_m(e_n)]$

In what follows, I will choose a slightly different, though equivalent, setting. I will speak of an arbitrary *valuation* $||$ as any function taking expressions to intensions appropriate to their category. Such valuations may very well be totally crazy: what we typically do is to restrict attention to the well-behaved ones. I say that $||$ *embeds* a lexical interpretation scheme if that scheme coincides with the restriction of $||$ to lexical elements. I call a valuation $||$ *compositional* if the following general relation holds:

$$|e(e_1, \ldots, e_n)| = |e|[|e_1|, \ldots, |e_n|]$$

We then let models contain compositional valuations, rather than lexical interpretation schemes.[51]

This finishes the description of the syntax and semantics of our 'single indexed pure categorial language'. To summarize:

1. Our language consists of infinitely many categories of expressions, described via a categorial grammar. $S$ and $N$ are basic categories. If $C$ and $C_1, \ldots, C_n$ are categories (basic or derived), let $C/C_1 \ldots C_n$ be a derived category.

2. We have a set of *indices*, $I$ and a domain of *objects $O$*, a set of *truth-values* $\{T, F\}$, and a valuation $||$.

3. The valuation $||$ assigns intensions to each expression appropriate to its category. *Appropriate intensions* for each category are functions whose type reflects the build up of the category, in the fashion described above.

4. Call a valuation *compositional* iff the semantic projection rule of function-application is met. Suppose we have an expression $e$ of category $C/C_1 \ldots C_n$, and expressions $e_i$ of categories $C_i$ (so that the complex expression $e(e_1 \ldots e_n)$ will be of category $C$). Then it must be that:
$$|e(e_1 \ldots e_n)| = |e|(|e_1|, \ldots, |e_n|)$$

**The permutation argument within single-indexed general semantics**

Let us introduce two further notions. First, call two compositional interpretations *sententially equivalent* if they assign the same intensions to everything of category $S$. Second, say that a valuation embeds the reference scheme $r$ (relative to index $i$) iff whenever $e$ is of category $N$, $|e|(i) = r(e)$. The aim now is to show that we can always construct *sententially equivalent* compositional valuations which embed arbitrarily permuted reference schemes.

The theorem we will prove shows that any coherent[52] assignment of semantic values (i.e. appropriate intensions) to sentences can be made compatible with arbitrarily permuted reference schemes. We do this by defining the notion of a $\phi$-variant of a valuation, where $\phi$ is a permutation of the domain of objects. If $||$ is a valuation embedding at index $i$ the reference scheme $r$ that, e.g. assigns Susan to "Susan"; then the $\phi$-variant $||_\phi$ of $||$ will embed at $c$ a reference-scheme $r_\phi$ that assigns $\phi(\text{Susan})$ to "Susan". Rather than setting up a permuted lexical interpretation scheme and proving by induction that the valuations determined by the original and $\phi$-variants appropriately match, I shall define directly the $\phi$-variant valuations, in a way that makes obvious the fact that they are sententially equivalent to the valuation from which one starts, and that they embed an appropriate permuted reference scheme. The challenge will then be to show that the definition is legitimate: i.e. that the permuted valuation, so-defined, is compositional. We must show that the semantic values that it assigns to complex expressions arise in the appropriate way from the semantic values it assigns to the lexical basis.

The construction of the $\phi$-variant valuations is as follows:

- Let $\phi$ be an arbitrary permutation of $O$. Define recursively the $\phi$-image of an intension as follows:[53]

    - If $f$ is an appropriate intension for $S$, then $f^\phi = f$
    - If $f$ is an appropriate intension for $N$, then $f^\phi = \phi \circ f$
    - If $f$ is an appropriate intension for $C/C_1 \ldots C_n$,
      then $f^\phi : r^\phi \mapsto g^\phi$ iff $f : r \mapsto s$.

- Given a valuation $||$, let the $\phi$-permuted valuation $||_\phi$ assign to $e$ its $\phi$-image; i.e. for each $e$, $|e|_\phi := f^\phi$ where $f = |e|$.

Notice that, by construction, a $\phi$-variant of a valuation $v$ will embed a reference scheme that is permuted by $\phi$. Moreover, a valuation and its $\phi$-variant will be sententially equivalent, again by construction. What remains to be shown is that the $\phi$-variant is a legitimate valuation of the language—that is, that it is compositional in the sense delimited above. That isn't hard:

**Theorem 1.** *For any permutation $\phi$, $||_\phi$ is compositional if $||$ is.*

*Proof.* Suppose $||$ is compositional. Take an expression $e$ of category $C/C_1 \ldots C_n$, and expressions $e_1 \ldots e_n$ of categories $C_1 \ldots C_n$ respectively. The complex expression $e(e_1 \ldots e_n)$ will be of category $C$). By compositionality of $||$, we have:

$$|e(e_1 \ldots e_n)| = |e|(|e_1|, \ldots, |e_n|)$$

We need to show that:

$$|e(e_1 \ldots e_n)|_\phi = |e|_\phi(|e_1|_\phi, \ldots, |e_n|_\phi)$$

We show this for the case $n = 1$, the other cases being trivial extensions. Write $e_1 = d$. From the compositionality of $||$ we have: $|e(d)| = |e|(|d|)$; i.e. $|e| : |d| \mapsto |e(d)|$ By the definition of the $\phi$-image this gives us: $|e|^\phi : |d|^\phi \mapsto |e(d)|^\phi$. Given the way that $||_\phi$ is defined, we can rewrite this as: $|e|_\phi : |d|_\phi \mapsto |e(d)|_\phi$, i.e.:

$$|e(d)|_\phi = |e|_\phi(|d|_\phi)$$

so compositionality is secured. $\qquad\square$

**The permutation argument within double-indexed general semantics**

Indexical expressions are commonly thought to require the semantic values we have been discussing to be changeable: an expression may express different such values when uttered in different contexts. One option within a general semantics is to include context (represented as a 'centred world') as one index within the above account.[54] Within that setting, our proof runs just as before.

Alternatively, one can separate world, time, delineations etc. from context, to "double index" the general semantics.[55] Syntactically, everything is unchanged. Within the semantics, however, expressions are assigned *characters* rather than intensions, where a character is a function that maps contexts to Lewisian compositional intensions. We need the notion of an 'appropriate' character for a given category of expression. Therefore, say that a character is of type $\tau$ iff it maps contexts to intensions of type $\tau$; and wherever the appropriate intension for category $c$ was type $\tau$, appropriate characters will be those of the same type.

Formally, characters resemble 'Carnapian intensions' rather than the 'compositional intensions' of Lewis' general semantics. Consider, for example, an intransitive verb (category $S/N$).

In Lewis' general semantics, its semantic value was the function from semantic values of type $N$ to semantic values of type $S$. On the double-indexed proposal, by contrast, its semantic value is a function from contexts to intensions; i.e. the context index is treated in a 'Carnapian' way. The upshot of this is that the semantic projection rule cannot simply be function-application, as within the original Lewis treatment. We can still have a single rule appropriate to arbitrary concatenations, however. Consider an expression $a(b)$, where $b$ is of category $c_1$ and $a$ is of category $c_2/c_1$. If $a$ has character $A$ and $b$ has character $B$, then the character of $a(b)$ maps context $\alpha$ to the intension that results from "compositionally" applying $A(\alpha)$ to $B(\alpha)$.

We need to spell out what the permuted interpretation function is for this new setting. This is defined in the obvious way. If $|e|$ is $\sigma$, where $\sigma : c \mapsto g$, then $|e|_\phi$ is $\sigma^\phi$, where $\sigma^\phi : c \mapsto g^\phi$, where $g^\phi$ is the $\phi$-image of the intension $g$. Notice that since $g^\phi = g$ when $e$ is a sentence we have that $|e|_\phi = |e|$ for $e$ a sentence. The permuted interpretation not only holds the intensions of sentences invariant, it holds the characters invariant also.

**Derivative invariance results**

We have seen that we can leave the semantic values of sentences invariant under permutations of the reference-scheme in a rich class of settings. The ultimate aim of such results is to argue for extreme inscrutability, by saying that all such 'sententially equivalent' semantic theories are equally good at fitting the data which the interpretationist provides. Such data may not take the form of a pairing of sentences with something as rich as the compositional intensions or characters that we have just been looking. It might rather involve a pairing between a sentences and propositions—the set of circumstances in which the sentence is to be true.

In such a setting, what we need for our purpose is that semantic theories that differ by containing permuted reference-schemes can assign to sentences the same truth-conditions (function from possible worlds to truth-values). This follows immediately from the above result once we note that the truth-conditions of a sentence are determined by the semantic value of that sentence in the rich settings just canvassed. If semantic theories are sententially equivalent, then they will assign to sentences the same truth-conditions.

To illustrate this, let us show how to construct two senses of 'truth-conditions' for a sentence $S$. These are what are known within 'two-dimensional modal logic' as the $C$-intension and the $A$-intension of a sentence.[56]

For $w$ a centred world (i.e. a context), define the matching-relation $m(x, y)$ to hold between world $w$ and string of indices $i$ iff the world, time etc. components of $i$ match those given by $w$. First, we will define the $C$-intension of a sentence $s$ at a context $c$. This is a set of centred worlds given by:
$$w \in C(s) \Longleftrightarrow \forall i(m(w, i) \rightarrow (|s|(c) : i \mapsto T))$$

Second, we will define the $A$-intension of a sentence. This is the set of centred worlds given by:

$$w \in A(s) \Longleftrightarrow \forall i(m(w, i) \rightarrow (|s|(w) : i \mapsto T))$$

To see the difference between these two constructs, consider the sentence "he is Beckham", uttered while pointing at the famous footballer. The $C$-intension is the set of worlds where the person indicated (i.e. Beckham himself) is identical to Beckham. On standard assumptions, then, this is the necessary proposition—the set of all possible worlds.[57] The $A$-intension is the set of worlds where *the person indicated by the pointing* is Beckham. It is false, therefore, with respect to possible worlds where Beckham stayed at home and sent out his double to make the public appearance. Notice that the $A$-intension is contingent where the $C$-intension is necessary. It is arguable, in the light of these distinctions, that the data for a semantic theory might be given

in the form of a pairing of sentences with $A$-intensions.[58] It is reassuring for the inscrutabilist, then, that all these sentential constructions are invariant under the permuted valuations. This can be easily checked by noting that the only appeal to the valuation $||$ in the above definitions is in application to *sentences*, where, as we have seen, $||$ and $||_\phi$ coincide.

# B  Indexical-inscrutability: a cut-and-shunt argument

We have examined arguments focused on what might loosely by termed "synchronic" inscrutability of reference.[59] We now prove a "diachronic" version. An expression such as "London" will, on the standard interpretation, designate London no matter what the context is.[60] Likewise, the deviant interpretations assign to that term a single object—Sydney, say—at all contexts. Because the reference of a term is at all times the image under the permutation of its original reference, the pattern of indexical dependence of reference will be the same on each of the permuted interpretation schemes.

We now have the resources to extend the arguments so that they imply a kind of 'indexical inscrutability'. The *extreme indexical inscrutability* thesis is that, if $c$ and $c'$ are distinct contexts, then there is no fact of the matter whether $e$ designates the same object at $c$ as at $c'$. To get this result, we 'cut-and-shunt' some of the permuted interpretations constructed above. The intuitive case is straightforward. By the results of §2, we have two valuations that always agree with each other on the content assigned to whole sentences, no matter what context is chosen. If we construct a third valuation, which matches valuation-1 at some contexts, and valuation-2 at the remainder, it will also have this property—all three will agree on the content of sentences at each context. Our first valuation might have it that 'London' denotes London at all contexts, whereas our second valuation has it that it denotes Paris at all contexts. The constructed valuation will agree with the first on some contexts, with the second on others—so it depicts 'London' as indexical. Yet it is sententially equivalent to the originals, and so under the current assumptions, the interpretationist has no grounds for ruling it out. What we do below is sketch how this intuitive idea could be formalized within the double-indexed semantic setting given above.

Take the candidate double-indexed valuation, $||^2$. Relative to a given context $c$, the designated valuation determines what would be a valuation within the *single*-indexed semantics, i.e. an assignment of compositional intensions to expressions:

$$|e|^{\langle 1, c \rangle} := |e|^2(c)$$

We have seen how to construct sententially equivalent $\phi$-variant interpretations $||_\phi$ of a arbitrary single-indexed valuation $||$ in the proof of the permutation result for the single-indexed case, and we can apply these techniques here, yielding the $\phi$-variants $||_\phi^{\langle 1, c \rangle}$.

Now assign permutations $\phi_i$ to contexts via a function $\mathbb{P}$. Relative to $\mathbb{P}$, we construct a new double-indexed valuation, which we call the $\mathbb{P}$-variant of $||^2$ and denote $||_\mathbb{P}^2$.

Given an expression $e$ and a context $c$, we let $||_\mathbb{P}^2$ assign to an expression $e$ the intension given by $||_\phi^{\langle 1, c \rangle}$ at $e$, where $\phi = \mathbb{P}(c)$, i.e.:

$$|e|_\mathbb{P}^2(c) = f \iff |e|_{\mathbb{P}(c)}^{\langle 1, c \rangle} = f$$

We now argue that $||_\mathbb{P}^2$ is sententially equivalent to $||^2$. At any given context $c$, the intension assigned to a term is that assigned to it on some $||_\phi^{\langle 1, c \rangle}$. However, the intension assigned by $||_\phi^{\langle 1, c \rangle}$

to a sentence is exactly that assigned to it by $||^{\langle 1,c \rangle}$, by our previous results. By construction $||^{\langle 1,c \rangle} = ||^2(c)$, so for all contexts $c$, $||^2(c) = ||^2_{\mathbb{P}}(c)$. QED.

By choosing $\mathbb{P}$ appropriately, we get radical indexical overgeneration. Consider "London", which we shall suppose to be of category $N$. Given contexts $c$ and $c'$ pick $\mathbb{P}$ where $\mathbb{P}(c)$(London)=London and $\mathbb{P}(c')$(London)=Paris. Then on the $\mathbb{P}$-variant interpretation, "London" will pick out London as uttered at $c$ and Paris as uttered at $c'$. Since the variant interpretation is sententially equivalent to the original interpretation, given BEST=FIT, it is not determinately incorrect.

Equally, by choosing the contexts and permutations carefully, we could take an indexical expression and give it a constant reference. One could choose $\mathbb{P}$ so that the permutation $\phi$ appropriate to a given context $c$ maps the speaker of that context to London. Relative to that valuation, 'I' is a non-indexical expression referring to London irrespective of context. Again, by the argument above, such interpretations are not determinately excluded.

# C    Comments on McGee on permutation arguments.

In a recent paper, Vann McGee (2005a) has argued that permutation arguments for extreme inscrutability break down when we try to apply them to the modal case. Arguing that this sort of thing does not happen was the central claim of the appendices above, where I presented a argument that there were what I called "$\phi$-permuted variants" of a given general semantic theory that were sententially equivalent: they assigned the same semantic values—and so, in particular, the same truth conditions—to sentences.

The setting presupposed above has a *single* domain of objects from which the extensions assigned to names at worlds are drawn. We have said nothing about what the *domain* of each world is. Some are prepared to defend the single-domained interpretation of modality as the *right* one. But even if we give this up, nothing in the proof would change if we added into the model a function from worlds to subsets of this domain which represents 'what exists at the world'. It is true that, as McGee notes, the permuted interpretations will typically end up assigning relative to world $w$, extensions to names that feature objects that don't exist at $w$. But whether or not this is a problem for the inscrutabilist seems a *philosophical* rather than *technical* issue: formally everything is smooth. It is argued in the text above that the inscrutabilist should have no qualms about accepting the resulting 'inscrutability of existence' result.

McGee's other argument concerns a different sort of setting for the permutation argument: one where the permutations are allowed to vary independent from context to context (just as, in the cut-and-shunt argument above, the permutations are allowed to vary independently from world to world).[61]

One of McGee's points is that when we run permutation arguments in this setting, we typically turn rigid intensions into rigid intensions: a short argument for this is given in the text above. But of course there is no technical *incoherence* in assigning non-rigid intensions to names. So if the argument is to stop here, then it must be because of some non-technical objection to non-rigid intensions.

McGee thinks there is something technically amiss with this argument, however. He notes that the usual clauses for variables in modal logic, e.g.

'$Fx$' is true at $w$ on variable assignment $v$ iff $v(x)$ is within the extension of '$F$' at $w$

in effect makes variables *automatically* rigid. They are assigned an object *once and for all* by the variable assignment, which they 'carry round' from world to world. McGee exploits this

standard treatment of variables to make trouble for this form of the permutation argument.[62]

The trouble here emerges from the (standard) syncategorematic axiom governing the semantics for variables. There are numerous alternative ways for the inscrutabilist to give a semantic treatment of variables, which will allow her to avoid McGee's objections.

A Lewisian treatment of variables sees them as expressions of category $N$, and accordingly, receiving an $N$-intension on a given interpretation. 'Variable assignments' (construed as functions from variables to objects) become just one more index, and the *intended* interpretation satisfies the following:

the extension of '$x$' at index $i$ and variable assignment $v = v(x)$

This makes variables under the *intended* interpretation rigid (since they pick out $v(x)$ irrespective of what values the other indices take). But it leaves open the possibility of *unintended* or *permuted* interpretations where the variable-intension is non-rigid. And indeed, simply following through the recipe for assigned name-intensions to expressions of category $N$, will typically make the variable intension non-rigid. McGee's argument lapses.

If one doesn't like assigning intensions to variables, another place for the inscrutabilist to take a stand is to hold that her permuted semantics *incorporates different syncategorematic axioms*. Suppose $\mathbb{P}$ is the function from indices to permutations which is the basis for the permuted interpretation of a language (so that the extension of 'Peter' at $i$ is $\mathbb{P}(i)$(Peter). Then the appropriate axiom for the inscrutabilist to appeal to is:

'$Fx$' is true at $i$ on variable assignment $v$ iff $\mathbb{P}(i)(v(x))$ is within the extension of '$F$' at $w$

Again, where is the argument that such 'permuted' syncategorematic axioms are *not* part of the 'intended' semantic theory?

Lastly, there are ways of giving semantics for quantified languages that do not involve saying anything about free variables. One such setting is used by Wright and Hale, in the appendix to their (1997), where they prove that permutation arguments of the form that McGee is interested in go through for a language of quantified modal logic. The interested reader is directed to their work for details.

# Notes

[1]My working assumption is that a type theoretic, multiply intensional semantics of the form described by Lewis (1970) suffices to interpret natural language. For discussion of how permutation arguments look within Davidsonian semantic theory, see Author (2008).

[2]Of course, the correlation may not be perfect: sometimes people speak falsely! Different theorists deal with imperfections in the correlation in different ways. It is the basis for the famous defense of the 'principle of charity' in Davidson (1973): the default assumption must be that people speak truly. Lewis (1975) takes a different approach, with the relevant correlation being between the states of the world that speakers *believe to obtain*, and the sentences they utter.

The approach just described contrasts sharply with what Davidson (1977) calls the 'building block' approach where ones philosophy of linguistic representation gives priority to subsentential reference. A *locus classicus* here is Field (1972).

[3]A variation would be to look at patterns of assent and dissent among *ideal* speakers. This is the form discussed in Putnam (1980)

[4]Lewis in later work complicates this picture by putting extra constraints on successful semantic theory. I discuss this briefly below.

[5]Versions of the permutation argument are given in Jeffrey (1964); Quine (1964); Field (1975); Wallace (1977); Putnam (1978); Davidson (1979); Putnam (1981). The most detailed presentation of the results that I know of is in the appendix to Hale and Wright (1997).

[6]That in the appendix to Hale and Wright (1997).

[7]Thanks here to NN and a referee for this journal, for urging me to consider the applicability of my considerations to moderate inscrutability theses.

[8]These, and other objections, are discussed in more detailed in my Author (2005, ch 6.)

[9]For the view that understanding is knowledge of meaning, see (Dummett, 1976). For the view that it at least consists in *beliefs* about meaning (Heck, 2005). For the view it consists in some other kind of attitude (Garcia-Carpintero, 2000; Stanley, 2002).

[10]One might find tension if competence consisted in knowing the meaning of sentences alone, where 'meaning' is richer than simply Lewisian 'truth-conditions'; for example, if it consists in knowing what *structured proposition* the sentence expresses, as canvassed by Soames (1989). I will not discuss this here, however.

[11]To emphasize, McGee's arguments were originally intended to show that the permutation arguments themselves do not go through. But I think they are better seen in the current light. To anticipate, his arguments that permutation arguments fail require us to *build it in* to the semantic setting we choose, that e.g. the extension of something at a world must be exist at that world; or that certain expressions cannot be but rigid (the strongest case for the latter being in the case of *variables*, which in most possible worlds semantics are treated as automatically rigid). As shown in the appendix, there's no formal reason for the inscrutabilist to accept such settings

(including the putative rigidity of variables). It seems better, therefore, to present the case as one where we are invited to consider such features *costs* that the inscrutabilist incurs

[12]I give detailed consideration to each of the above in my Author (2005, ch 6.)

[13]For examples of this tactic, see McGee and McLaughlin (1994) on 'pluth', and the view ascribed to Hartry Field in Loewer (2005) and McGee (2005b).

[14]Talking of different 'interpretation functions' is, of course, just to talk about certain mathematical objects, functions from words to objects.

[15]Compare Hale and Wright (1997).

[16]Consider, for example, Williamson (2000), who thinks that at worlds where I was never born, I nevertheless exist as an *abstractum*. On a Williamsonian view, the most we can get out of McGee-style considerations is that it is indeterminate whether or not we are referring, at a world, to things that are *concrete* at that world. That result lacks the shock value of the original thesis.

[17]The rigidity of variables is even more deeply embedded, but I am not aware of any *arguments* for this thesis, so I'm taking it that the simple availability of workable systems where variables are non-rigid, as sketched in the appendixes, suffices here.

[18]This, of course, is not meant to be a criticism of Kripke's arguments: his main target was the kind of non-rigidity that descriptivists postulated; in that dialectical context, he was entitled to ignore possible extreme inscrutability.

[19] A striking disanalogy between the temporal and modal cases is that we are *causally* connected to (say) Beethoven, whereas we are not causally connected to mere possibilia. But for this disanalogy to have relevance, we must explain why causal contact is a precondition on referring. This would be straightforward on, say, a metasemantics that proposes to identify the reference relation with certain causal relations. But the inscrutabilist is working within a setting where the metasemantics is not causally constrained in this way (otherwise the argument for extreme inscrutability couldn't even get started). So again, it's not clear what the argument against the inscrutabilist is supposed to be.

[20]The case is very different from that of *empty names*: empty names are those that *fail to secure* a referent. What is under consideration here is a case where reference is *secured* to something that fails to (actually) exist.

[21]In the literature on supervaluationist vagueness, analogous coordinations are used to explain so-called 'penumbral connections': $A \wedge \neg A$ can be determinately false, even if $A$ and $\neg A$ are each indeterminate.

The fallacious move is of the following form: from $\neg Dp$, $\neg Dq$; move to $\neg D(p \wedge q)$. This is directly comparable to the fallacious move within modal logic (which likewise has some initial appeal): moving from $\Diamond p, \Diamond q$ to $\Diamond(p \wedge q)$.

[22]In an appendix to this paper, I give the formal details.

[23]This sketch of the extreme indexical inscrutability result is proved in formal detail for rich languages in the appendix to this paper.

[24]Indeed, to hold the thesis it is at best indeterminate whether the token arguments we give are logically good seems dangerously *self-undermining*: for the grounds on which we accept this conclusion is exactly on the basis of logical argumentation. So it looks like we might get some kind of pragmatic paradox in this situation: having to hold a thesis while admitting one's grounds for holding it are not determinately good.

The denial that certain token arguments have a special epistemic status in virtue of their logically valid form seems non-negotiable to me. But I'm not sure that the accusation that denying this would be *undermining* in this way. After all, to deny that this or that argument is in *logical* good standing is not to deny that they have other desirable epistemic features (e.g. that the premises entail the conclusion). And one wishing to go on biting bullets at this point might claim that it is no embarrassment to be relying on arguments that have this status. Connectedly, some may already be inclined to resist the first premise, on the basis of an independent commitment to the idea that the conclusion of an argument following logically from the premises is *epistemically irrelevant*. This strikes me as wildly implausible, but for excellent discussion and advocacy of something close to this, see Harman (1988, ch.1.). Thanks to Nick Zangwill for discussion here.

[25]One might try to resist the argument by arguing that, against appearances, token arguments involve only a single context. Williamson suggests something like this view: see XXX on p.NN.

[26]The independent case for this turns on the correct interpretationist response to scenarios such as Strawson's 'massive reduplication' worlds (Strawson, 1959, ch.1). I briefly discuss these in Author (forthcoming). See Hawthorne (forthcoming) for some critical discussion of the 'eligible characters' proposal.

[27]I do not mean to use 'logical consequence' in an overly narrow manner, however. I am quite amenable to thinking that 'analytical' inferences (say, from '$M$ is red' to '$M$ is coloured') are in the relevant sense valid, though they are clearly not valid in the narrowly logical sense. The semantic treatment of validity is easily extendable to such cases.

[28]Hawthorne (op cit) makes a number of sensible further points. Competent deduction may require more than inferring in a way that accords with a valid form: it might require that we be aware of this fact, or have followed a proof. I have no quarrel with this at all: indeed, my claim here is, in effect, that the stability of the language involved must form one part of competent deduction.

Hawthorne further claims: "a competent deduction need not consist of a formally valid argument" (p.35). I take it that by this he means that there is no reason to restrict closure to cases of *narrowly logical deduction*. But in saying this he in no ways denies that narrow logical deduction is one of the central and important cases where closure holds. Further, I contend that competent deduction in general always involves arguments of types which are semantically valid *in a broad sense*: for semantic characterization of validity only coincides with narrowly logical validity on one particular choice of what to hold invariant across reinterpretations. If we allow 'meaning postulates' to constrain legitimate reinterpretations, then the cases Hawthorne mentions: e.g. 'that is red' to 'that is coloured' will indeed be semantically valid. But this broad notion of validity, again applying in the first instance to inference-types, raises exactly the same issues as the narrow notion.

[29]The notion extends to arbitrary sets of formulae taken as premisses.

[30]Note that in some presentations, the admissibility constraints are obscured, since the semantic theory contains explicit axioms fixing the interpretation of certain expressions. In many presentations of the model theory of first order or modal logic, the semantics of logical connectives is handled by including explicit axioms governing these expressions within the model theory. In a 'general semantic' framework (cf. Lewis, 1970) logical constants are assigned particular semantic values, and the need for explicit decisions over what should count as an admissible re-interpretation of the language is more evident.

[31]Published as Kaplan (1989), but in circulation in manuscript form since the 1970's.

[32]For present purposes, I ignore other options, such as taking dated *utterances* to be truth-bearers.

[33]One might try to distinguish the *context of utterance* and *context of interpretation*. See below.

[34]Relativization to possible worlds is only one of the ways in which truth-predicates can usefully be parameterized. Others include relativization to variable assignments, delineations, to times and locations. See Lewis (1970, 1980) for general discussion.

[35]It is tempting to gloss this as 'necessary truth preservation' under any admissible re-interpretation. This is misleading, as the usual characterization allows arbitrary *frames* (i.e. re-interpretation of the 'worlds' component of the designated semantics.)

[36]Strictly speaking, we should generalize over models, not just interpretations, where a model is a tuple that specifies, in addition to the interpretation, a 'frame' for the interpretation: sets of points playing the role of context, etc. and any needed accessibility relation on this space.

This way of characterizing validity is 'local' in character, in the sense of Williamson (1994). An alternative to the above would be the 'global' style there characterized, whereby each model for the language contains additional structure corresponding to 'privileged points': which in the designated model correspond to the actual world (for the modal case), the context of utterance (for the indexical case), etc.

[37]I take a context to be a centred world, in the way suggested by Chalmers (1996). See Lewis (1980) for discussion of the utility of centred worlds within a (generalized form of) Kaplanian indexical semantics.

[38]This treatment of validity I regard as the basic idea of the logic of demonstratives. The distinctive theses associated with that work—i.e. that 'I am here now' is a logical truth, and that the necessitation rule fails for logical validity, arise from other features. The first of these is the way that Kaplan restricts admissible interpretations to those that hold fixed the 'character' of expressions (he does this implicitly, by using specific axioms to fix the reference of basic indexical expressions). The second special feature is the way that Kaplan combines indexical and intentional logics to define LD-validity. Neither of these special features will be presupposed here.

[39]On a language-of-thought model of reasoning, movements in thought from one set of judgements to another fall within this compass.

[40]As presented, the argument assumes that all the premisses are uttered in a single context. Clearly the argument will generalize to situations where the premisses are to be evaluated with

respect to distinct contexts within a invariant set. Similar considerations will also cover cases where different parts of a single sentence are to be evaluated with respect to distinct contexts.

[41]Lewis (1980) adopts something like this view, based on, among other things, the consideration just canvassed.

[42]The basic ingredients for this story are the account of inference involving demonstratives outlined in Campbell (1994), and its generalization to a wider class of indexical expressions by Prosser (2005). To convert the 'Fregean' setting to one amenable to Kaplanian treatment, we can appeal to an account of 'dynamic words' based on the treatment of proper names in Kaplan (1990).

[43]Davidson's truth-theoretic semantic framework, and the structured-proposition theory of Soames (1989), give rise to a finer grained notion of 'truth-conditions'. I argue elsewhere (reference omitted) that the Davidsonian setting prevents us from providing a permutation style proof of overgeneration; and a similar situation arises with the structured proposition framework. The inscrutability problems are merely relocated, however. Developing a theory of how one pairs sentences with fine-grained truth-conditions in such settings is correspondingly harder.

[44]In fact, it is controversial whether ordinary proper names are in category $N$.

[45]I keep the bracketing for notational convenience, though it plays no part of the official definition. It would be a trivial matter to alter the definition of well-formedness to allow for parentheses, but it would be a distraction.

[46]Montague used a richer set of rules *within* the general semantics itself to make the formal language closer to English syntax. Cf. (cf. Partee, 1996, §3), Dowty (1979, ch.1) for discussion.

[47]Lewis (1970) includes category $C$ (common nouns) as another basic category, having as appropriate intension a function from indices to sets of objects. For the sake of simplicity, I ignore this here.

[48]Note that there is a single 'type' containing all objects—when we define intensional operators over this framework, we will end up with a 'single domain' semantics. Plausibly, we want to be able to say that there are things which do exist, which might not exist had some other situation been actual. The solution is to pair each world-index $w$ with a partition of the type of objects into two: those that exist at $w$, and those that do not. We then interpret the 'ordinary' loaded existential quantifier relative to $w$ be restricted to those objects that exist at $w$. Cf. §**??**, below.

[49]This is formulated for derived categories with two elements only. For 'binary branching' languages we need no more. For the more general case, we need some additional notation to mark functions from sequences of types to types. (McCawley, 1980, ch.13)

[50]The compositional intensions which Lewis favours differ from the 'Carnapian intensions' that form the basic framework for Montague (1970). To characterize a Carnapian intension, one first defines the notion of an appropriate *extension* in a way paralleling the recursive formulation above. The extension for a name will be an individual of type $o$; the extension for a sentence will be a truth-value of type $t$, and in general an appropriate extension for an expression of category $c/d$ will be of type $\langle \alpha, \beta \rangle$, where $\alpha, \beta$ are the types appropriate to $c, d$ respectively. An appropriate *Carnapian intension* for an expression of category $Q/R$ will be a function from

indices to appropriate extensions, i.e. of type $\langle i, \langle \alpha, \beta \rangle \rangle$ where $\langle \alpha, \beta \rangle$ is the appropriate extension for category $Q/R$. The deficits of a framework based on Carnapian intensions—centrally, its inability to deal with 'intensional' expressions such as the predicate 'is rising'—are discussed in Lewis (1970) and are addressed by Montague by assigning certain expressions semantic values which are not of the Carnapian kind. See Thomason (1974b).

[51]The two settings are equivalent. Given a model that supplies a lexical interpretation $||$, the corresponding compositional valuation is the valuation $v_m$ determined by $||$. Given a model that supplies a compositional valuation, the restriction of that valuation to atomic expressions is the corresponding lexical interpretation. A minor advantage of this setting is that even semantic projection rules are no longer required within the semantic theory, minimizing 'syncategorematic' elements that are unmotivated in a metasemantic setting.

[52]Call an assignment of semantic values to sentences *coherent* if there is at least one compositional valuation that embeds that assignment.

[53]If, following Lewis (1970), we included an extra basic category $C$ ("common nouns") mapping indices to sets of objects, then the permuted variant of $|e|$ for $e$ in category $C$ should be $|e| \circ f^1$.

[54]See Lewis (1970, appendix).

[55]Consider, for example, what happens if context changes through the utterance of a sentence say (cf. Lewis, 1979, p.241). So long as we can assume that there's a unique context associated with each lexical element, context will determine an intension for each such element, which then collectively determine an intension for the whole. The story would be far more complex if we proceed in terms of functions from context-index pairs to truth-values.

[56]For the terminology, see Jackson (1998). For alternative labels ('secondary intension' and 'primary intension' respectively), see Chalmers (1996). The $C$-intensions also correspond to what Stalnaker (1978) calls the 'horizontal proposition' associated with a sentence. Note that $A$-intensions/primary-intensions are not the diagonal propositions of Stalnaker (1978). Stalnaker's diagonal propositions are not semantic constructs, but are defined pragmatically in terms of agents' common knowledge of the semantic content of their words.

[57]I set aside concerns about worlds where the individual does not exist.

[58]Unpublished work by MM discusses this issue.

[59]The terminology would be strictly appropriate if the only variable aspect of context were temporal. Clearly there are many non-temporal ways in which contexts can vary.

[60]This follows from the orthodox treatment of terms as having 'constant character', following Kaplan (1989). Some, such as Jackson (1998), claim that the names are indexical. They maintain that, at the least, they can vary depending on what world they are uttered in.

[61]In fact, this is not even formulable in the 'compositional' setting we are here, where the intensions for derived categories are functions from intensions to intensions: rather, it presupposes a 'Carnapian' setting where the intensions for an expression in a derived category is a function from indices to 'extensions' appropriate to that category. Thus, instead of the intensions of predicates mapping name intensions to sentence intensions, they will map worlds to (functions from name extensions to sentence extensions).

[62]For a taste of the problems, consider the following, with respect to a deviant interpretation that at some other possible world puts the man Peter within the extension of 'is a raspberry':

> There is something that is Peter and which is possibly a raspberry

The assignment of Peter (the object) to the variable in this sentence will make true both the first conjunct (since that object is picked out by 'Peter' at the actual world) and the second (since the object falls under 'is a raspberry' at some possible world). But we can suppose that this quantified modal sentence is *false* on the intended interpretation.

# References

Ajdukiewicz, K. (1935). 'Die syntaktische konnexität'. *Studia Philosophica*, **1**, 1–27. English translation: 'Syntactic Connection', in McCall S.(ed.), *Polish Logic 1920-1939* (Oxford University Press: Oxford, 1967), 207-231.

Author (2005). 'The inscrutability of reference'. Ph.D. Thesis, University of NN.

Author (2006a). 'Eligibility and inscrutability'. *Philosophical Review*.

Author (2008). 'Permutations and Foster problems: two puzzles or one?'. *Ratio*, **XXI**(1).

Author (ms. 2006b). 'Gavagai again'.

Campbell, J. (1994). *Past, Space and Self.* MIT Press, Cambridge, MA.

Campbell, J. (1997). 'Precis of *Past, Space and Self* and replies to critics'. *Philosophy and Phenomenological Research*, **57**(3), 633–634, 655–670.

Chalmers, D. (1996). *The Conscious Mind.* Oxford University Press, Oxford.

Cresswell, M. (1973). *Logics and Languages.* Methuen, London.

Davidson, D. (1973). 'Radical interpretation'. *Dialectica*, **27**, 313–28. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.125–140.

Davidson, D. (1974). 'Belief and the basis of meaning'. *Synthese*, **27**, 309–23. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.141–154.

Davidson, D. (1977). 'Reality without reference'. *Dialectica*, **31**, 247–53. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.215–226.

Davidson, D. (1979). 'The inscrutability of reference'. *The Southwestern Journal of Philosophy*, pages 7–19. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.227–242.

Dowty, D. R. (1979). *Word Meaning and Montague Grammar.* D Reidel, Dordrecht, Holland.

Dummett, M. A. E. (1976). 'What is a theory of meaning? Part II'. In G. Evans and J. McDowell, editors, *Truth and Meaning: Essays in semantics*, pages 67–137. Oxford University Press, Oxford.

Field, H. H. (1972). 'Tarski's theory of truth'. *Journal of Philosophy*, **69**, 347–375. Reprinted in Field, *Truth and the Absence of Fact* (Oxford University Press, 2001) pp. 3-29.

Field, H. H. (1975). 'Conventionalism and instrumentalism in semantics'. *Noûs*, **9**, 375–405.

Garcia-Carpintero, M. (2000). 'A presuppositional account of reference-fixing'. *Journal of Philosophy*, **109-147**, 71–104.

Hale, B. and Wright, C. (1997). 'Putnam's model-theoretic argument against metaphysical realism'. In C. Wright and B. Hale, editors, *A Companion to the Philosophy of Language*, pages 427–457. Blackwell, Oxford.

Harman, G. (1988). *Change in view: principles of reasoning*. Bradford Books/MIT Press, Cambridge, MA.

Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford University Press, Oxford.

Hawthorne, J. (forthcoming). 'Craziness and metasemantics'. *Philosophical Review*.

Heck, R. (forthcoming 2005). 'Reason and language'. In C. MacDonald, editor, *McDowell and his Critics*.

Jackson, F. (1998). *From Metaphysics to Ethics: A defence of conceptual analysis*. Oxford University Press, Oxford.

Jeffrey, R. (1964). 'Review of *Logic, Methodology and the Philosophy of Science*, ed. E. Nagel, P. Suppes and A. Tarski'. *Journal of Philosophy*, **61**, 79–88.

Kaplan, D. (1989). 'Demonstratives'. In J. Almog, J. Perry, and H. Wettstein, editors, *Themes from Kaplan*, chapter 17, pages 481–563. Oxford University Press, New York; Oxford.

Kaplan, D. (1990). 'Words'. *Proceedings of the Aristotelian Society*, **supp LXIV**.

Lewis, D. K. (1970). 'General semantics'. *Synthese*, **22**, 18–67. Reprinted with postscript in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 189–229.

Lewis, D. K. (1975). 'Language and languages'. In *Minnesota Studies in the Philosophy of Science*, volume VII, pages 3–35. University of Minnesota Press. Reprinted in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 163-88.

Lewis, D. K. (1979). 'Scorekeeping in a language game'. *Journal of Philosophical Logic*, **8**, 339–59. Reprinted in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 233-49.

Lewis, D. K. (1980). 'Index, context and content'. In S. Kanger and S. Öhman, editors, *Philosophy and Grammar*, pages 79–100. Reidel, Dordrecht. Reprinted in Lewis, *Papers on Philosophical Logic* (Cambridge University Press, 1998) 21–44.

Lewis, D. K. (1983). 'New work for a theory of universals'. *Australasian Journal of Philosophy*, **61**, 343–377. Reprinted in Lewis, *Papers on Metaphysics and Epistemology* (Cambridge University Press, 1999) 8–55.

Lewis, D. K. (1984). 'Putnam's paradox'. *Australasian Journal of Philosophy*, **62**(3), 221–36. Reprinted in Lewis, *Papers on Metaphysics and Epistemology* (Cambridge University Press, 1999) 56–77.

Loewer, B. (2005). 'On Field's *Truth and the Absence of Fact*: Comment'. *Philosophical Studies*, **124**, 59–70.

McCawley, J. (1980). *Everything that Linguistics have always wanted to Know about Logic (but were ashamed to ask)*. University of Chicago Press, Chicago.

McGee, V. (2005a). "Inscrutability and its discontents". *Noûs*, **39**, 397–425.

McGee, V. (2005b). 'Two conceptions of truth?: Comment'. *Philosophical Studies*, **124**, 71–104.

McGee, V. and McLaughlin, B. (1994). 'Distinctions without a difference'. *Southern Journal of Philosophy*, **supp XXXII**, 203–251.

Montague, R. (1970). 'Universal grammar'. In R. Thomason, editor, *Formal Philosophy: Selected papers of Richard Montague*. Yale University Press, New Haven and London.

Partee, B. H. (1996). 'The development of formal semantics in linguistic theory'. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 11–38. Blackwell, Oxford.

Prosser, S. (forthcoming 2005). 'Cognitive dynamics and indexicals'. *Mind and Language*.

Putnam, H. (1978). 'Realism and reason'. In *Meaning and the moral Sciences*, pages 123–140. Routledge and Kegan Paul, London.

Putnam, H. (1980). 'Models and reality'. *The Journal of Symbolic Logic*, **45**(3), 421–444. Reprinted in Benacerraf and Putnam (eds.) *Philosophy of Mathematics: Selected readings*, second edition (Cambridge University Press, Cambridge: 1983).

Putnam, H. (1981). *Reason, Truth and History*. Cambridge University Press, Cambridge.

Quine, W. V. (1950). *Methods of Logic*. Holt, New York.

Quine, W. V. (1960). *Word and Object*. MIT Press, Cambridge, Mass.

Quine, W. V. (1964). 'Ontological reduction and the world of numbers'. *Journal of Philosophy*, **61**. Reprinted with substantial changes in Quine, *The Ways of Paradox and Other Essays: Revised and enlarged edition* (Harvard University Press, Cambridge, MA and London, 1976) pp.212—220.

Salmon, N. (1998). 'Nonexistence'. *Noûs*, **32**, 277–319.

Soames, S. (1989). 'Semantics and semantic competence'. *Philosophical Perspectives*, **3**, 575–596.

Stalnaker, R. (1978). 'Assertion'. *Syntax and Semantics*, **9**, 315–332. Reprinted in Stalnaker, *Context and Content* (Oxford University Press, Oxford, 1999) pp.78-95.

Stanley, J. (2002). 'Modality and what is said'. *Philosophical Perspectives*, **16**, 321–344.

Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics*. Methuen, London.

Thomason, R., editor (1974a). *Formal Philosophy: Selected papers of Richard Montague*. Yale University Press, New Haven and London.

Thomason, R. (1974b). 'Introduction'. In R. Thomason, editor, *Formal Philosophy: Selected papers of Richard Montague*. Yale University Press, New Haven, CT.

Unger, P. (1980). 'The problem of the many'. *Midwest Studies in Philosophy*, **5**, 411–67.

Wallace, J. (1977). 'Only in the context of a sentence do words have any meaning'. In P. French and T.E. Uehling, Jr., editors, *Midwest Studies in Philosophy 2: Studies in the Philosophy of Language*. University of Minnesota Press, Morris.

Williamson, T. (1994). *Vagueness*. Routledge, London.

Williamson, T. (1997). 'Sense, validity and context'. *Philosophy and Phenomenological Research*, **57**(3), 649–654.

Williamson, T. (2000). 'Existence and contingency'. *Aristotelian Society sup. vol*, **73**, 181–203. with printer's errors corrected, *Proceedings of the Aristotelian Society* 100, 1 (2000): 117-139.