

# There's No *Me* in Meta—Or Is There?

Don Perlis

Department of Computer Science  
University of Maryland  
College Park, MD 20742  
perils@cs.umd.edu

## Abstract

Despite the egocentric-sounding title, this is not about myself but rather about the concept of self and how it may relate to that of meta, in the context of various considerations, including formal and informal notions of self-reference, reference in general, theories of consciousness, and the active logic approach to commonsense reasoning.

## Introduction

Meta-this or meta-that carry the sense of an X (a this or a that) which is being examined from a vantage point that allows X to be taken in as a whole--or at least in large chunks that can be used in characterizing general properties of X.

- *Metalinguage* is the mode of expression used in examination of a mode of expression, i.e., of (possibly another) language, e.g., in asking questions such as: what does that word mean? what language are they speaking? is that a valid expression? how is that spelled? can you repeat that?
- *Metamathematics* is an examination of mathematics (or a portion thereof). It confronts questions such as: is it consistent? what rules govern it? and so on.
- *Metacognition* is the examining of (one's own) cognition. One comes to characterize some of one's mental processes: my memory failed me; I am no good at arithmetic; I easily remember faces; I have a good sense of direction.

Several themes percolate through these examples, just below the surface. One is that meta-X involves a person—a reasoner—who is doing the examining of X. Another is that X itself is a kind of reasoning people do: we do math, we cognize, we use language. Yet another is that in doing meta-X, we are examining something about ourselves,

namely our doing of X. And a fourth is that one takes a step back from performing X in order to examine X, i.e., to perform meta-X.

[Metaphysics does not quite fit this description; instead of being an examination of the practice of finding laws of nature, it examines (among other things) why there are laws of nature at all, what are “reality” and “existence”, etc. But for that reason, metaphysics is not a case of meta-X at all. In fact, the term seems to have originated in reference to that topic Aristotle’s writings that came after his writings on physics, thus effectively eliminating metaphysics from our concern here. My thanks to Aaron Sloman for calling this to my attention.]

So it would appear that in performing meta-X we are stepping back from our X-performance in order to examine it; we stop doing X so we can look at it. This resonates with the familiar dictum: we cannot catch ourselves in the act--but only in the aftermath of ceasing that act.

To quote from William James (1892) at some length, both for context and for the pure joy of Jamesian prose:

[Substantive and Transitive States of Mind]

Let us call the resting-places the 'substantive parts,' and the places of flight the 'transitive parts,' of the stream of thought. It then appears that our thinking tends at all times towards some other substantive part than the one from which it has just been dislodged. And we may say that the main use of the transitive parts is to lead us from one substantive conclusion to another.

Now it is very difficult, introspectively, to see the transitive parts for what they really are. If they are but flights to a conclusion, stopping them to look at them before the conclusion is reached is really annihilating them. Whilst if we wait till the conclusion be reached, it so exceeds them in vigor and stability that it quite eclipses and swallows them up in its glare. Let anyone try to cut a thought across

in the middle and get a look at its section, and he will see how difficult the introspective observation of the transitive tracts is. The rush of the thought is so headlong that it almost always brings us up at the conclusion before we can rest it. Or if our purpose is nimble enough and we do arrest it, it ceases forthwith to itself. As a snowflake crystal caught in the warm hand is no longer a crystal but a drop, so, instead of catching the feeling of relation moving to its term, we find we have caught some substantive thing, usually the last word we were pronouncing, statically taken, and with its function, tendency, and particular meaning in the sentence quite evaporated. The attempt at introspective analysis in these cases is in fact like seizing a spinning top to catch its motion, or trying to turn up the gas quickly enough to see how the darkness looks....

Yet in metamathematics it turns out that (for example via Gödel numbering) a good deal of the meta-X (X being mathematics) is itself part of X. To be sure, the concerns or aims of meta-X are not quite the same as those of X; but there is a large overlap of the two, at least in this case. And the typical metalanguage for a natural language—say English, for example—is English itself.

So, here is the theme/poser of this essay: when we examine our performance of a cognitive activity X, must that examining (itself an activity) effectively halt our X-performance, or at least so change it that it no longer is what we were attempting to examine? And if so, can it then be examined in pristine only from without, by a highly distinct process? Or can one perform a mental activity that continues full-blast (“in a headlong rush”, to borrow from James) even *while* being looked at from within—perhaps even an activity that *is* that very looking at its own performance of itself? Answers tend to involve one of two notions which we may refer to as *hierarchical* (looker is separate from lookee) or *loopy* (looker and lookee can be intertwined and even one and the same). These terms will be further explained as we proceed.

However, it seems clear that the notion of a self, or a me, is near and dear to the theme at hand. And now I hope the meaning of the title is somewhat clearer as well.

Consideration of our theme will lead us to a number of traditionally far-flung topics: informal (natural-language based) self-reference; formal self-reference in mathematical logic; the problem of reference in general; consciousness; commonsense reasoning; and mistakes.

We will now take these up in turn.

## Informal self-reference

Perhaps the most famous example of a self-referential utterance is the so-called *Liar*:

L: This sentence is false.

The Liar, or L for short, has two curious features: (i) it appears to refer to itself, and (ii) it appears to contradict itself. The latter feature is the one that has received most attention, but we will focus instead on the former. At this early stage let us simply note that there is an issue as to what, if anything, can guarantee that the word “This” in L succeeds in referring to L itself, as opposed (say) to some other sentence that may have recently been uttered or pointed to. Presumably it is our agreement that it so refer; but then there is personal agency involved.

Consider these two sentences that directly speak to personal agency:

I am using the letter “L” to refer to the sentence to the right of that letter.

The above sentence calls attention specifically to how an expression is to be taken as referring, according to the agent-speaker.

Here I am speaking in English in Chicago as I scratch my head, wondering how I will complete this sentence that I will complete...now.

In this sentence there is no explicit mention of reference or meaning, but words such as “Here” and “I” must be understood as referring to the present time and to the speaker. Moreover, surely this sentence is true—or was true when it was uttered—and surely in uttering it I referred to myself and to that very sentence-uttering process as it was occurring. The process refers to itself, wearing, so to speak, its meaning on its sleeve, and it achieves that in virtue of my decision to make it do that and of your understanding that this is what I am doing.

Perhaps more precisely, I am referring to my uttering activity as it occurs, and that referring action is simply that very uttering activity itself. The uttering activity is a self-referring process. Not the sentence, but the activity. What activity? Not mere production of words. Uttering in this context is supposed to mean something like: attempting to convey something to someone, and the self-referential uttering activity is one that is (or involves) an attempt to convey itself—i.e., its own self-conveyingness—to someone.

But this then rests upon a clear reference for “I”, presumably once again an agent with referential intentions, a matter seemingly far removed from the usual concerns in discussions of the Liar.

Pretty weird stuff. Self-referring entities tend to be suspect; yet instead of being accidental oddities that creep in because the expressive power of our languages (formal or otherwise) is too lax, they are essential. In fact, the very possibility of reference of any kind supervenes on a special kind of self-reference, that I have called strong self-reference, that will come up again later.

Indeed, as Grice (1957) has urged, every utterance is implicitly of this sort, a self-commenting or self-meta-self, as if it were of the form:

With this utterance I am attempting to convey the meaning of this utterance to you.

Well, that might not work, it's not clear it has a lot of meaning. But the Chicago utterance-process above has a clear meaning. It is also clear that it is an attempt to convey something, and that that something is that very process itself.

One natural-enough reaction to the above circumstances is to suppose that some sentences (such as the Liar) are, after all, neither true nor false, indeed neither true nor not true. This may seem to play fast and loose with the word "not"--after all, "not" simply asserts the failure of what follows it. But perhaps there is something hidden here. In order to be a candidate for failure in the matter of its truth, any sentence must first have a potential truth that can fail, i.e., it must have a clear enough meaning that can be measured against some criterion of truth. How does a sentence acquire a meaning? This is the subject of much dispute, and we will return to it a bit later on.

## Formal self-reference

The examples we have considered so far are informal, based largely on commonsense notions. However, it is not hard to capture similar behaviors in more formal dress. A key component of the formalization is the Diagonal Lemma, which asserts that in any reasonably expressive formal theory F, for each unary wff  $Px$  there is a sentence  $p$  such that in F it is provable that  $p \leftrightarrow \neg Pp$ . Given the Diagonal Lemma various formal results along the lines of the Liar follow.

With more work (Godel) one can devise a wff  $Thm$  of F such that, for all sentences  $s$ ,  $Thm s$  is provable in F if and only if  $s$  is provable in F; that is,  $Thm$  is a provability-predicate for F.

Now from the Diagonal Lemma, there is a sentence  $g$  (a so-called *Godel sentence*) such that in F it is provable that  $g \leftrightarrow \neg Thm g$ , i.e., that  $g$  is equivalent to its own unprovability in F. It follows that if  $g$  is provable, then so

is  $\neg Thm g$ , and hence we get Godel's Theorem: either  $g$  is unprovable in F (and thus true, in the sense that what it "asserts"--its own unprovability---holds), or F is inconsistent (since if  $g$  is provable then so are both  $Thm g$  and  $\neg Thm g$ .)

But such formal results are purely syntactic, and reference (let alone self-reference) plays no real role. For example, that L refers to anything is irrelevant to the proof that the formal version of the Liar is inconsistent. And the Godel sentence  $g$  does not really refer to anything at all, let alone to its own unprovability.

And who cares? Isn't self-reference just a curiosity?---an accident arising as a side-effect of an over-expressive language, with surprisingly useful but equally accidental application in formal logic, and of no deep significance in itself? Cannot we then simply ignore the bad (contradictory) cases and welcome the good? As it turns out, we cannot: self-reference, far from being an unimportant side-effect, is central to reference, hence to meaning, and arguably to meta-X as well. This leads us to set out two highly dissimilar approaches to meaning assignment.

[I should point out that there has been a tremendous amount of work done on paradoxes of self-reference, formal and informal. For example, Gilmore (1974) and Kripke (1975) in particular have formulated two closely related approaches that not only defuse the paradoxical (contradictory) aspects but also retain (much of) the self-referential aspects. The focus of such work in general has been on clarifying the notion of truth, which is slightly tangential to our concerns here.]

## Hierarchical or Loopy?

Since the above formal results are just that, formal (syntactic) and not dependent on semantics, then it might be possible to keep the advantages of certain "seeming" self-reference (as in Godel's Theorem) without the disadvantages of the Liar (such as inconsistency). Tarski showed how to do this by means of a restriction on how languages refer. He posited a hierarchy of languages  $L_1, L_2, \dots$  where each  $L_{(j+1)}$  has expressions that refer only to objects in a previously defined language  $L_j$ . There is then no expression that can refer to itself. This approach simply banishes self-reference from expression altogether, while leaving intact the syntactic vestiges needed for the Diagonal Lemma (and useful formal results).

On the other hand, this hierarchal approach seems to banish too much. There are perfectly innocuous but semantically-based cases of self-reference, such as:

This sentence has five words.

Yet this is not expressible in the Tarski Hierarchy. Nor is the following pair of straightforward sentences, each referring to the other (one happens to be false):

The sentence below has seven words.  
The sentence above has six words.

Yet if these are clear enough, if a sentence can clearly refer to another specific sentence, above or below itself, then why not to itself?

The sentence right at this spot is in English.

For that matter, in referring to something above itself, a sentence implicitly refers to its own position to give meaning to “above” or “below”. Since the meanings of some sentences are not captured hierarchically, we will borrow a phrase of Hofstadter’s (2007), and say they have *loopy semantics*.

How does meta-X bear on hierarchical or loopy semantics? The hierarchical case is simple enough: each language  $L_{(j+1)}$  is a meta-language for the previous one, it takes a step back, provides the vantage point, for commenting on other sentences. There is no *me* in these sentences since they always refer to another domain, never to the one where they sit.

Loopy semantics, on the other hand, is precisely that of sentences that self-refer, whether directly, or indirectly via (a loop through) other sentences such as in the pair above. But is this then not also simple, straightforward? Yes and no. Here is where we start to see some complexities in the concept meaning.

The problem, at its core, is that the hierarchical semantics does not really address reference at all. In postulating that language  $L_{(j+1)}$  is “about” objects (e.g., sentences) in  $L_j$ , one is simply making a stipulation, not explicating what it is for a sentence to mean anything, or refer. It is only by means of an agreement among whichever logicians happen to be participating in the discussion, that an expression refers to anything at all.

The deictic “this” of natural language (as in “This sentence is false.”) has been by-passed altogether in formal treatments. Indeed, reference (or semantics) of any kind is traditionally placed outside a formal language, as a function defined on expressions in the language, mapping to an external domain. The language in question does not typically have an expression that stands in for this function; and even if it did, what would determine that standing-in relation? It is as if meaning, or truth, is always one step removed, leaning on some agreement lying outside whatever language is used.

So, as far as the hierarchical approach goes, a map between symbols and referents is arbitrary, leaning on a decision to use that map, and not some other, by an agent who intends to use that map. Contrary to what Putnam (1975) has claimed, meaning *is* (at least partly, and very significantly) in the head (of the agent using that meaning). Only when this issue is faced head on, do we encounter genuine cases of reference, and the possibility of genuinely self-referring expressions. And these turn out to be precisely the loopy cases.

## General reference

A lesson we draw is that self-reference proper has largely been left untouched by the very large literature purportedly on that subject. This is because reference has largely been left untouched, or rather pushed to the sidelines, via Godel numbering or a similar artificial mechanism that leans on external agreements to bring reference in at all. Attention has focused, rather, on formal counterparts of self-reference that do, to be sure, carry with them a substantial potential for contradictoriness, in close analogy to their informal—but more genuinely self-referential—sources. But while this attention has produced much of great importance, it has left much out as well. First and foremost is this problem: can there be representation (meaning, reference) without an agent who chooses to so represent? And secondarily, what is the relation between reference in general, and self-reference? Third, what can be said about “genuine” informal self-referential expressions, in light of answers to the former questions? Space does not permit a detailed discussion, but my answers are: no representation without intentional stipulative agency, reference supervenes on self-reference, and genuine or “strong” self-reference is a special “loopy” agency bordering on consciousness.

## Consciousness

Wait a minute! Consciousness? Aren’t we aimed at understanding meta-X here?

Well, a me, a self, arguably is the essence of consciousness (however, this is incredibly controversial). But if so, and if a me or a self is part of meta, then meta involves consciousness. But we have seen two versions of meta: hierarchical and loopy. The former is little more than a sequence; the latter is mysterious. Hierarchical meta-X presumably is akin to a Tarskian pair of levels, the object level X and the monitoring level meta-X. There seems little more to be said about it, except that someone sets up a map by which the expressions at the meta-X level refer to items at the X level. But then whoever sets up this map is really

the determiner of meanings, without whom there is no particular designated map, and X and supposed meta-X have no particular relationship. And meaning determiners, as far as we know, are always people, or at least agents with intentions.

We are faced then once again with agents: agents that perform activities and also refer to those activities (some of which are self-referring).

Consider again that English is its own metalanguage. We typically talk about English in English, and for the most part do not need a special set of meta-English terms to do this, beyond ones such as “word” and “means” and “spell” and “sentence”. But even here, such terms refer to bits of English only because at least one human has taken them to so refer.

What is it, then, for an agent to “take” one thing to “refer” to another? Consider a primitive case: coining an expression, explicitly linking a symbol *s* to a referent *r*. This would seem to be no more nor less than an intention to use *s* as a stand-in for *r* in certain contexts. Following this trail, we now ask what it is to intend something, and we are smack dab in the middle of both philosophy of language and philosophy of mind. And to reinvoke Grice, every utterance is a case not merely of intending, but also of intending listeners to understand that the utterer intends that intending. Can all this happen in the absence of a fairly sophisticated (and quite possibly conscious) cognitive engine? Moreover, the natural languages that we use for expression of intentions are—as noted—their own metalanguages, allowing loopy self-reference made possible by our intentions to so refer: we speak of ourselves, not just past or future, but our immediate present self and present activity including the activity of noting that activity.

So, once again, does meta have a me? If meta involves reference, and if reference involves agency with intentions, including intentional self-referring activity, and if that in turn is at least a hint of a self, then yes.

Thus a bare-bones agent self-reference may be the most basic kind of reference. What is bare-bones self-reference like? Imagine yourself stripped little by little of this sensation, that thought, until all that is left is your own grasp of being there, a bare loopiness without—for the moment—any further trappings: no personal history, no connections with or knowledge of entities in the world other than the one thing: you-as-bare-awareness.

Such a minimal subjective state—if there is such a thing—I have dubbed the “ur-quale”: the most primitive sensation or feeling possible, namely that of simply being a being, an entity whose one activity is self-monitoringness (Perlis 1997, 2000).

This may be what is suggested in Piet Hein's poem *Evening and Morning Son—About falling asleep and waking up* (1972) [my thanks to Torkil Heiede for bringing this poem to my attention]:

The world disappears,  
a loop running smaller, until  
the thread is drawn out,  
and the space it encloses is nil.

Newborn of nothing,  
reluctantly starting to be,  
fumbling awareness awakens  
and finds that it's me.

Let us step back a little from such tenuous speculations, to a more engineering perspective, following an idea of John Perry (1979). He describes pushing his shopping cart along in an attempt to find the shopper whose cart is leaving a trail of sugar on the floor, only later to realize *he* is that shopper. But what is it that he learns? Perry explores this question at some length, for although on the one hand it seems obvious, on the other it is devilishly hard to say what it is. Still, it seems to matter, and not only for egocentric reasons. Our everyday actions seem tied to it.

## Commonsense reasoning

We can ask a related question about robot design. Consider a robot that can decide *it* is the robot who is leaking oil, upon hearing that robot #17 is leaking oil. What is it for robot #17 to know that it, itself, is that robot? How does this affect its behavior? Presumably it is quite important to have such a capability, e.g., for survival. See (Anderson and Perlis 2005) for more elaborate discussion of this idea.

It is worth stopping to ask what good such a thing might be (whether full-blast or loopy). One answer seems easy to come by: the self-examining that consciousness appears to provide has survival value. And in fact, much of recent work in AI has been aimed at providing useful self-examining capacities to automated systems, including commonsense reasoners. Indeed, that is presumably what this workshop is about. I turn then to an approach central to my own work, that not only seems to promise some usefulness for AI systems, but also bears a bit on some of the other issues we have been exploring. Namely, I will briefly discuss active logic.

Active logic is a type of “commonsense formal inference engine”; that is, it consists of a language and rules of

inference, but with a twist: the rules are sensitive to the actual physical passage of time. Thus the inference of  $Q$  from  $P$  and  $P \leftarrow \rightarrow Q$  is sanctioned only if that inference occurs at a time  $t+1$  when both  $P$  and  $P \leftarrow \rightarrow Q$  had been inferred at time  $t$ . In effect, time values are part of the language; but that in itself is no news: temporal logics have such as well. Active logic however ties time values to actual physical time-passage, by keeping track of the evolving current time. The active logic sentence  $Now(t)$  has the obvious meaning. But what is striking is the so-called “clock” rule: from  $Now(t)$  infer  $Now(t+1)$ . So we have a sentence which—if believe—leads to its own disbelief. The reason of course is that such a belief is time-sensitive, and as time passes the belief becomes outdated. But it has major consequences for what active logic can do. For just one example, it facilitates a smooth handling of contradictions. On the other hand, semantics becomes much more complex; see (Anderson et al, 2008).

I bring this up because active logic also provides a useful way to talk about differences, or the lack of differences, between metareasoning and object reasoning since the two are not distinguished at all in active logic. Each inference step involves looking back at the previous step to see what was inferred then, and on that basis drawing inferences at the new current step. Whether such an inference is modus ponens, based on the presence of  $P$  and  $P \leftarrow \rightarrow Q$  at the previous step, or instead is based on the presence of  $P$  and  $\neg P$  and yields  $Contradiction(t)$  (recording the fact that there was a direct contradiction at the previous step), makes no difference at all to the machinery. It is only to us that these inferences seem to exist at different levels, and then only if we are sensitized to X vs. meta-X as a hierarchical distinction from a young age in a formal logic class.

There is much more to be said, but we shall leave it and turn instead to the practical matter of mistakes.

## Meta and mistakes

As a technical branch of AI, commonsense reasoning is the subject of a great deal of study. Yet (as with many such things) there is no general agreement as to what it is, at least in sharp definitional terms.

Here is my own definition, offered as a kind of hypothesis.

Commonsense reasoning (CSR) is the form of metareasoning that monitors an activity for mistakes and then deals with them, sparing the main activity the embarrassment of making a fool of (or destroying) itself.

More specifically:

- (i) CSR consists of a module that *notes* mismatches between observation and expectation in a system’s performance, *assesses* any such, and *guides* a response into place.
- (ii) This is the essence of human common sense. We have expectations, note deviations from them, and decide on a response.
- (iii) Such a CSR module need not know much at all about the system in question, other than having access to at least some of the system’s pre- and post- conditions (expectations) for its actions and at least some of its observations (e.g., sensor readings).
- (iv) This module can be fairly simple, based on a core set of general kinds of things that can go wrong and general kinds of fixes for them. For instance: noise can interfere with sensor readings, in which case one can use another sensor or take repeated readings, or replace the sensor; data can be contradictory, and one can distrust one or more of the contradictands, seek corroborating evidence for them, change its expectations, give up and work on another problem, ask for help, etc.
- (v) The actual carrying out of such a response is not the CSR module’s job; it merely recommends, and the system must then be equipped for the repair.
- (vi) One particularly interesting case is that of a failure due to lack of some skill, so that training is needed. Training then can be recommended by the CSR module, and if training commences, the module monitors that, assesses its progress, and recommends when to stop training (whether due to its not working, or to its having succeeded).
- (vii) If such a module were to be built and put to use with a given AI system, that system would become far less brittle, and vastly better at dealing with anomalies, than any AI systems at present.
- (viii) Such a module could be general purpose, not built for any specific system or domain. This is because humans are not specific in that sense. We manage to muddle through in a wide variety of unanticipated changes.

We are in fact hard at work designing just such a module—as an outgrowth of active logic—which we call the *metacognitive loop*; see (Anderson et al 2006).

## Conclusion

What then can we say about our poser: is there is or is there not a *me* in meta? Well, to the extent that common garden variety so-called self-monitoring systems are “meta”, the answer seems to be a clear no. One part of a system can monitor another without any self-reference at all, as if the two parts were in effect distinct systems, one watching the other.

In the case of active logic’s form of meta-X, the answer appears to be a weaker no. While there is (so far!) no sharp notion of me or self in active logic, still an active logic sentence might refer to itself by invoking a description of itself (e.g., *the first sentence inferred at time t*). And there can be loops, with one sentence referring to a later one yet to come, which in turn may refer back to it.

But note that reasoning about a past or future or even present, does not in itself constitute an activity that reasons about its very self. And even in the case of the sentence italicized above, the self-description is akin to Perry’s describing himself in a third-person manner as “the person who...” without having the special form of self-knowledge indicated by use of the pronoun “me”.

Even the metacognitive loop (MCL) in the form presented above does not appear to embrace anything close to self-reference.

On the other hand, if MCL itself were to make mistakes and need to catch them, things might get more self-referentially interesting, especially if the mistake-noting were in some sense part of the mistake. That might nudge MCL (or active logic) toward more explicit (strong) self-reference.

How can such a thing be, and what could it possibly mean? Consider this:

*This activity is taking too long, it must be stopped,  
including this very reasoning about it, so other things  
get done.*

We have not yet managed (or even tried very hard) to produce such behavior. But it does seem that there is a computational advantage to it: such a system would be able not only to peer in on *other* processes and reason about and control them; it would also be able to do so regarding the *very same peering process that it is performing at the moment*, so that it can get out of its own way, so to speak, deciding to stop what it is doing—including that very deciding—and move on to something else.

It is time to move on—and so with this utterance I end my talk.

## References

- Anderson, M., Gomaa, W., Grant, J., and Perlis, D. 2008. *Active logic semantics for a single agent in a static world*. Artificial Intelligence. To appear.
- Anderson, M., Oates, T., Chong, W., and Perlis, D. 2006. *The metacognitive loop I: Enhancing reinforcement learning with metacognitive monitoring and control for improved perturbation tolerance*. Journal of Experimental and Theoretical Artificial Intelligence. 18(3): 387–411.
- Anderson, M. and Perlis, D. 2005. *The roots of self-awareness*. Phenomenology and the Cognitive Sciences. 4(3): 297–333.
- Gilmore, P. 174. *The consistency of partial set theory without extensionality*. In T.Jech, ed. *Axiomatic Set Theory*, 147–153. American Mathematical Society.
- Grice, P. 1957. *Meaning*. Philosophical Review 66:377-88.
- Hein, P. 1972. *Grooks, IV*. Doubleday, New York.
- Hofstadter, D. 2007. *I Am a Strange Loop*. Basic Books.
- James, W. 1892. *The stream of consciousness*. Psychology, chapter XI. Cleveland & New York, World.
- Kripke, S. 1975. *Outline of a theory of truth*. Journal of Philosophy 72:690–716.
- Perlis, D. 1997. *Consciousness as self-function*. Journal of Consciousness Studies.
- Perlis, D. 2000. *What does it take to refer?* Journal of Consciousness Studies.
- Perry, J. 1979. *The problem of the essential indexical*. Nous, 13:3–21.
- Putnam, H. 1975. *The meaning of meaning*. Mind, Language and Reality. Cambridge Univ. Press.