Thought, thoughts, and deflationism

Vann McGee

Abstract  Deflationists about truth embrace the positive thesis that the notion of truth is useful as a logical device, for such purposes as blanket endorsement, and the negative thesis that the notion doesn’t have any legitimate applications beyond its logical uses, so it cannot play a significant theoretical role in scientific inquiry or causal explanation. Focusing on Christopher Hill as exemplary deflationist, the present paper takes issue with the negative thesis, arguing that, without making use of the notion of truth conditions, we have little hope for a scientific understanding of human speech, thought, and action. For the reference relation, the situation is different. Inscrutability arguments give reason to think that a more-than-deflationary theory of reference is unattainable. With respect to reference, deflationism is the only game in town.

Keywords  Truth · Truth conditions · Reference · Inscrutability of reference · Thought · Deflationsim · Substitutional quantification · Frege · Tarski · Quine · Horwich · Field · Hill

1 Quine’s disquotationalism

Quine (1986, ch. 1) taught us to think of the truth predicate as a device for disquotation. If we take a sentence, say,

Snow is white,

and we enclose it in quotation marks,
“Snow is white,”
we get a name of the sentence. Applying the predicate “true,”
“Snow is white” is true,
cancels the effect of the quotation operation, giving us a sentence equivalent, for
nearly all purposes, to the one we started with.
“Disquotationalism” about sentences has come to refer to a package of views that
includes the doctrine that the word “true” is used to cancel the effect of quotation
marks. One part of the package is the idea that we can use the word as a logical device,
one that vastly extends the expressive power of ordinary logic. Ordinarily, to form
the conjunction of a set of sentences, we have to list its members, placing “and”s between
them. If we have the notion of truth, it will be enough to be able to name the set, since
we can say, “All the members of $S$ are true.” Thus, even though I am in no position to
enumerate all the Pope’s pronouncements, I can endorse them all, saying “Everything
the Pope says is true.” Although I am certainly not capable of forming an infinite list, I
can achieve the effect of infinite conjunction, saying “All the theorems of Peano
Arithmetic are true.”

The other component of the disquotationalist package is the negative, “defla-
tionist” doctrine that the only use of the notion of truth is its employment as a
logical device. Why do we even need a notion of truth? Leading disquotationalist
Christopher Hill answers (2014b, p. 54), paraphrasing and endorsing Quine’s view,
“because it provides us with the ability to endorse a large set of thoughts without
having to endorse each member of the set individually, and with the ability of
endorse single thoughts without having to formulate the thoughts explicitly.” The
notion of truth doesn’t have a substantial theoretical role. Of course, semanticists are
free to use truth as a logical device, just as anyone else, but it’s not a notion that’s
suitable for use as a key theoretical term in a scientific theory.

This is a startling doctrine. It looks as if something semanticists do, continually
every day, is to try to understand the connections between sentences and their truth
conditions. The deflationist has to say either that they’re not doing what they seem
to be doing or that their efforts are badly misguided.

The negative thesis isn’t a doctrine that Quine avows explicitly, but it’s in the
background. Scientific semantics, for Quine, is the study of verbal behavior, and
verbal behavior doesn’t pin down truth conditions. Concepts suitable for scientific
semantics are determined from the outside, by behavior. It looks as if we, as
speakers of the language, can specify truth conditions from inside the language, by
taking the (T)-sentence,

“Snow is white” is true if and only if snow is white,
as a partial definition of “true.” The appearance is misleading. If the meaning of the
defining phrase “Snow is white” is only determined up to stimulus synonymy, the
meaning of the defined phrase “‘Snow is white’ is true,” is likewise indeterminate.
In semantics, as Quine (1960) envisages it, the notion of truth conditions doesn’t
have a significant theoretical role. The part we would have expected it to play is
played instead by stimulus meanings.
2 “That” removal

Hill has expanded Quine’s basic idea in useful and insightful ways. The results he has accumulated over the years have been gathered together with his work on other philosophical topics into a valuable collection, (2014a). One of the older papers, (1983), builds directly on Quine’s idea of disquotational truth as an attribute of sentences. I am oversimplifying. Quine applied disquotational truth to so-called eternal sentences (1960, pp. 193f), whose conditions of application don’t vary by context. An occasion sentence, like “It is snowing” isn’t either true or false, but true or false in a context.

The later papers move in an un-Quinean direction. Hill continues to refer to the conception he develops there as the view that truth is a device for disquotation, but a more accurate name would be truth as a device for “that” removal. Start with a sentence. By prefixing “the thought that” we get a name of a thought. Adding “is true” gets us a sentence that expresses a thought equivalent to the thought expressed by the sentence we started with.

The phrase “the thought that” is puzzling, at least to me. I think of a thought as something mental, something that is within a person’s mind. We talk about people in close agreement as having the same thought, but that surely doesn’t mean that there is some entity that is inside both their minds. We talk about two people having the same thought the way we talk about two people having the same microwave oven, when their microwaves are the same model. The idea of thoughts that somehow exist outside anyone’s mind seems bizarre and a little creepy, not unlike the idea of extraphysical microwave ovens or extramathematical numbers.

Yet there is the standard English phrase “the thought that snow is white.” Why the definite article? Hill has a thought about the color of snow. He thinks that snow is white. We don’t say that Hill’s thought that snow is white is true if and only if snow is white. We say that the thought that snow is white is true, in spite of the fact that you have your thoughts and he has his.

Frege (1892), perhaps as a manifestation of his abhorrence of psychologism about mathematics, took the “the thought that” locution at face value. Thoughts, on his view, are abstract entities in a timeless realm removed from the world of experience. Hill, according to Frege, thinks that snow is white because his mind has somehow latched onto a certain abstract entity, the thought that snow is white.

Frege’s doctrine has become philosophical orthodoxy, although, to avoid the jarring conclusion that thoughts aren’t mental, people have employed the word “proposition,” in a technical usage, where Frege uses “thought.” Hill uses both terms interchangeably.

We have the same double usage for “concept.” Concepts are components of thoughts, and just as we have two kinds of thoughts, mental and abstract, we have two kinds of concepts, mental and abstract. To complicate things further, Frege (1893 and 1903, §3) has a third usage. Fregean concepts come in various types. The simplest type of concept is a function taking objects to truth values. Strictly speaking, the identity relation only holds between objects, but there is an analogous relation for concepts, and, in Frege’s usage, we can never have two different
concepts that are true of precisely the same objects. If all and only the creatures with hearts have kidneys, then the Fregean concept *creature with a heart* will be the same as the concept *creature with kidneys*. The other two ways of using the word “concept” make finer distinctions, so that, since it’s perfectly conceivable that there should be a creature that has a heart but no kidneys, the abstract concept *creature with a heart* is different from the abstract concept *creature with kidneys*, and likewise, a particular individual’s concept of a creature with a heart will be different from that individual’s concept of a creature with kidneys. What Hill calls a concept, Frege refers to as the sense of a concept expression.

3 Why I am not a deflationist

A point of pride among disquotationalists is that their view avoids the need for contentious metaphysical commitments. It’s hard to see how the boast is justified for the “that”-removal version of disquotationalism, for the version presupposes the existence of abstract thoughts, and that is a heavy metaphysical commitment. Propositionalism is widely accepted among analytic philosophers—indeed, one might say that it’s the characteristic doctrine of analytic philosophy—but acceptance of the propositionalist consensus is by no means universal. Quine is a notable dissenter.

The thing that most bothered Quine about propositions was the squishiness of their individuation conditions, but there are other worries as well. I think that the dog Cherry likes to chase squirrels, which means that my mind has grasped the thought that Cherry likes to chase squirrels. But how can Cherry form a part of a timeless abstraction? Supposedly, the truth conditions of abstract thoughts are known a priori, but how could one know anything about Cherry a priori? At one time, it seemed like there was an answer. The name “Cherry” abbreviates a complex definite description formed from timeless concepts we can know about a priori. After Kripke (1972), this proposal no longer seems workable.

As a rule, it’s easy to tell concrete objects apart, because they occupy different locations. We can specify a concrete object by saying where it is. For abstract objects, things are not so simple. What specifies an abstract entity, if anything does, is its “position” within the relevant theory. Frequently, we talk about abstract entities in a way that presupposes that the entity is uniquely specified; we talk, for example, about “the” number of planets. In order for the entity to be uniquely specified, there has to be a theoretical role that is filled by it and it alone. But often there will be no such role. Starting with Benacerraf (1965), an entire generation of philosophers of mathematics have fretted over the fact that mathematical usage determines the referents of mathematical terms, at best, up to isomorphism.

We don’t have a Benacerraf problem for propositions. We can tell different propositions apart by their different truth conditions. It’s controversial whether we can find distinct propositions with the same truth conditions. Without getting into that issue, we can at least see that we’ll have no difficulty discerning the proposition that Cherry likes to chase squirrels clearly enough to distinguish it from the proposition that snow is white. The two propositions have different truth conditions.
This is good news for propositionalism, but bad news for deflationism. The role of truth conditions in individuating propositions has nothing to do with the use of truth as a logical device.

The fact that we use the word “thought” ambiguously, for ephemeral mental entities and for timeless abstractions, may lull us into overlooking a problem. What is the connection between my ephemeral thought that Cherry likes to chase squirrels and the abstract thought that Cherry likes to chase squirrels? To say that I grasp the abstract thought, like grabbing a bird in an aviary, isn’t helpful. We don’t have mental fingers with which to grab things. There is nothing like a causal connection between what’s going on in our minds and relations that obtain timelessly within the realm of thoughts. There is something that’s reasonable and natural to say: The mentally occurring thought and the abstract thought have the same truth conditions. This is, again, bad news for deflationism, since the use of truth conditions to connect the two kinds of thought is separate from the use of truth as a logical device. There is a similar question with a similar answer about connecting the sentence “Cherry likes to chase squirrels” with the proposition that Cherry likes to chase squirrels.

I am tempted to join Quine in rejecting propositionalism. There is certainly plenty of vacant space in the antipropositionalist campground. It’s not that I object to entities that are abstract and theoretical. It’s that I object to theoretical entities that aren’t pulling their weight, and it doesn’t seem to me that, once we relinquish the grasping metaphor, propositions are doing a lot of useful work.

A commonsense idea that philosophers have adopted enthusiastically is that a good way to understand why people do the things they do is to ask what the people want and why they believe that acting as they do will help them get the things they want. To know what their desires are, we have to understand how the world would need to change in order for their wants to be fulfilled. We can reformulate the question by looking for a proposition that expresses the content of the agent’s desires, then inquiring about the truth conditions of the proposition, but I don’t see how bringing in the propositional middleman improves our understanding. Desire is a state of the agent’s mind, and to understand why she does what she does, we need to know the fulfillment conditions of her mental state. Belief is likewise a state of the agent’s mind, and to understand why she does what she does, we need to ask what the world would have to be like for her beliefs to be true. Again, I don’t see how the propositional middleman earns his pay.

It seems to me that, when we’re not engaged in metaphysics, most of the things we want to say in terms of propositions can be said as well without them. As we shall see below, the tools Hill develops make it possible to follow out this idea systematically. Hill himself is a confirmed propositionalist, but the machinery he develops is versatile.

Folk psychological explanations of human behavior require us to determine the truth conditions of an agent’s beliefs and the satisfaction conditions of her desires, whether or not the determinations proceed by way of propositions. Allowing the truth conditions of beliefs and the satisfaction conditions of desires to assume a fundamental explanatory role requires the kind of substantive mental-state-to-world connections that deflationism disavows. We shouldn’t be content with folk psychology. We should look for explanations of how and why agents come to
have the beliefs and desires they have and how these mental states are embodied physiologically. We can’t predict with any confidence how closely the scientific psychology of the future will resemble the folk psychology of today, but it would be imprudent to rule out on philosophical grounds the possibility that truth conditions of belief and satisfaction conditions of desires will continue to play a key theoretical role.

The situation is similar for individual behavior and group behavior. Sounds issue from the boss’s mouth and, as a consequence, the workers load bricks on wheelbarrows and cart them to the north side of the building under construction. Their actions depend, in a systematic and predictable way, on the precise sounds emitted by the boss. Had the sound been slightly different, the workers would have taken the bricks to the south side of the building. To understand why the workers did what they did, we have to recognize that the boss was giving an order, and we have to recognize the conditions of satisfaction of the command. To identify the conditions of satisfaction of speech acts, it is necessary to be able to identify the truth conditions of assertions. Hill (2014c, p. 66) advocates the standard disquotationalist position that truth and related semantic notions don’t require scientific elucidation. But if science can’t account for why speech acts have the conditions of satisfaction they have, it can’t account for human verbal behavior. Communication by language is crucial to the organization of human societies, so that, if science can’t account for verbal behavior, it won’t be able to account very well for nonverbal behavior either. I hate to sound apocalyptic, but there is a danger a deflationary understanding risks putting both human social organization and human thought beyond the reach of scientific understanding.

As Hill puts it (2014c, p. 66), deflationism denies that truth, reference, and satisfaction “have a robust nature that can be elucidated by science, metaphysics, or normative inquiry.” In Leon Horsten’s (2011, p. 4) formulation, “truth is a concept without a nature or an essence”. I don’t know what would be required for something to have a robust nature. Why are “Truth is disquotation” or “Truth is a logical device” not answers to the question “What is the nature of truth?” on a par with “Truth is correspondence” and “Truth is the ideal limit of rational inquiry”?

Deflationism is sometimes explained negatively, by saying what deflationism rejects. It rejects the thesis that truth consists in the correspondence of a thought with a state of affairs that obtains. This explication doesn’t succeed. Hill (2014b, p. 57) shows that it is possible to give a deflationist reading of the notion of correspondence to an obtaining state of affairs.

A way of thinking about deflationism that I find easier to understand was expressed by Hartry Field (1986, 1994): The deflated notions shouldn’t play a significant role in causal explanations. This is a notion of deflationism that I find clear enough to reject. We find the notion of truth conditions continually in use in giving causal explanations, and the concept of truth is a component of the notion of truth condition.

I don’t know of any good reason to think the true sentences or the true thoughts have anything metaphysically significant in common. I am attracted to Lewis’s (1986) idea that, even though the actual world is dear to our hearts, there is nothing metaphysically special about it. It follows that there is nothing metaphysically
special about the propositions that are true in the actual world, i.e., the propositions that are true. The metaphysically significant question to ask about a proposition is “In what worlds is it true?” which is a version of the question, “What are its truth conditions?”

My own inclination is to think we should relinquish the negative, deflationist component of the disquotationalist package, and allow semantics to assume a reputable position within the sciences. Indeed, it appears that Field and Hill have sketched a promising program doing just that. We start out defining truth for our own language, or our own thoughts, disquotationally. The reason for taking that as our starting point is that that is the place that is most easily accessible epistemically. The methodology is straightforward: Start with what you know, then work your way toward things that are harder to know. Metaphysically, there is nothing special about me and my thoughts, but epistemically there is something special, for me: These are the thoughts that it’s easy for me to know about. Extend the notion of truth to other languages and other minds by learning how to translate those languages and thoughts into our own. If we already knew the truth conditions for the other language, we could impose the constraint that the translation preserve truth conditions, but that’s not our situation. Our situation is trying to discern truth conditions and correct translations simultaneously.

Where I disagree with Field and Hill is that I think that disquotation and translation provide us only the beginning of an adequate understanding of truth. An adequate scientific understanding requires us to go beyond merely listing truth conditions. It requires us to understand how and why particular expressions and particular mental states developed the truth conditions they have, as a part of a grand project of learning how matter learned to speak and think.

There are sure to be disruptions along the way. Vagueness is likely to cause problems. Hill (1987) examines an artificially simple language free of vagueness. Once we bring vagueness into the picture, we may well want to conclude that borderline attributions are such that neither they nor their negations are true. If “Clare is poor” is such a sentence, we’ll find that we can’t uphold the (T) sentences,

“Clare is poor” is true if and only if Clare is poor,

and

“Clare is not poor” is true in and only if Clare is not poor,

since together they entail:

Either “Clare is poor” or “Clare in not poor” is true.

Thus it appears likely that we’ll eventually have to abandon some of the (T)-sentences for our own language, even though we initially regarded them as our apodictic starting point; see McGee and McLaughlin (forthcoming). Scientific progress is often thus: Two steps forward, one step back.

The conclusion of my anti-deflationist diatribe is this: We shouldn’t abandon hope for scientific semantics on metaphysical grounds, even less on antimetaphysical grounds.
The central disquotationalist thesis is this: If we take an eternal sentence, either enclose it in parentheses or prefix “the thought that,” then add the words “is true,” the result will be a sentence equivalent to the sentence we started with. We can take “equivalence” here to be analytic equivalence, if we’re not squeamish about the notion of analyticity. Equivalence is ensured by the meaning of the word “true,” and it’s not dependent on a substantial linguistic or psychological inquiry. We’re stipulating that the word “true” is to be used in such a way as to ensure the material equivalence of “’Snow is white’ is true” and of “The thought that snow is white is true” with “Snow is white.”

Tarski (1935) took the derivability of the (T)-sentences as a condition of adequacy on a proposed definition of truth, and he showed how, in a wide range of cases, an adequate definition could be obtained by first defining the auxiliary notions of reference and satisfaction. The method requires logical resources in the metalanguage substantially beyond those available in the object language. In situations where we lack such resources, Tarski (1935, pp. 256ff) suggested taking the (T)-sentences themselves as axioms implicitly defining truth. It wasn’t a suggestion he was happy with, because the axioms lacked the power to establish even the most rudimentary generalizations. Thus we can prove for each sentence that it and its negation aren’t both true, but we can’t derive the general claim that a sentence and its negation are never both true.

Paul Horwich (2000) took up the proposal Tarski disdained, only talking about propositions in place of sentences. Tarski wanted to avoid metaphysical entanglements. Horwich, despite labeling himself a “deflationist,” has no such aim, developing an account that requires a massive investment in metaphysical heavy machinery. He postulates a function, designated “the thought that,” taking propositions to objects, a function “true” taking objects to propositions, and a function “iff” taking ordered pairs of propositions to propositions. He combines them to get a function taking a proposition $p$ to the proposition that the thought that $p$ is true iff $p$. The range of this function constitutes the axioms of Horwich’s theory. What a peculiar use of the term “axiom”! Ordinarily, we think of the axioms of a theory as theses adherents of the theory believe, but many of Horwich’s axioms are propositions human beings are incapable of even comprehending. There are thoughts human beings are incapable of grasping. The thoughts of dolphins are a presumptive example. For each dolphin thought, there is a corresponding (T)-thought, a curious amalgam of dolphin and human concepts, which, presumably, neither species is capable of understanding. Yet unless we can grasp all the (T)-thoughts, even the dolphin-based ones, we can’t be said to fully grasp the concept of truth. What Horwich has done, it seems to me, is to describe an axiomatic theory of truth, perhaps an axiomatic theory employable by superhuman intelligences. But he hasn’t given an axiomatic theory of truth.

The system Horwich describes is open to Tarski’s complaint. We can’t derive significant generalizations from the axioms. Horwich struggles to overcome this difficulty, but Hill (2014b) argues persuasively that his efforts are unsuccessful.
Hill proposes a way to overcome the difficulty, by a proposal structurally analogous to what Tarski did. Tarski showed that we could take a finitely axiomatized first-order theory implicitly defining satisfaction and convert it to an explicit second-order definition of satisfaction, from which he obtained an adequate explicit second-order definition of truth. From the explicit definition, the law of contradiction and the other desired generalizations readily follow. Hill similarly wants to convert the implicit definition provided by the (T)-thoughts into an explicit definition by enriching the logic, this time by using substitutional quantification:

\[ x \text{ is true } = \text{Def} (\Sigma p) ((x = \text{the thought that } p) \land p). \]

(This oversimplifies a bit and will be amended later.) There is much dispute about the individuation conditions for thoughts, but there is a minimal constraint nearly everyone will agree to:

\[(\Pi p) (\Pi q) ((\text{the thought that } p = \text{the thought that } q) \to (p \leftrightarrow q)),\]

This minimal constraint enables us to derive the (T)-thoughts. An apparently harmless “syntactic” condition,

\[(\Pi p) (\text{the negation of the thought that } p = \text{the thought that } \neg p),\]

lets us derive the general law that a thought is true if and only if its negation isn’t true. General laws describing the truth conditions for the other sentential calculus connectives are obtained in the same way.

Like Tarski’s account, Hill’s theory gives us a language-specific theory of truth. For Tarski, truth is a relative notion. A sentence is true in a language in a context. The same sentence, according to Tarski, can appear in different languages with different truth values. One can’t meaningfully ask whether a sentence is true unless one has an answer, possibly tacit, to the question “True in what language?” Propositional truth isn’t language dependent. I might express by asserting the sentence “Carlos es rico” the same thought you express by asserting “Carlos is rich.” The fact that the truth of a thought isn’t language- or context-relative is worrisome. It can happen that, in the morning, you use the sentence “Carlos is rich” to express a thought that’s true, and in the afternoon you use the very same sentence to assert a thought that’s false, where what’s changed in the meantime isn’t that Carlos has lost his money, but rather a change in the contextually determined comparison class. Yet morning and afternoon, you assert the thought that Carlos is rich. I complained earlier about taking the phrase “the thought that” at face value, on the grounds that, though we are both said to have the thought that Carlos is rich, your thought is in your head and mine in mine. We now see that the difficulty also appears intrapersonally.

The notion of propositional truth defined by way of the substitutional quantifier isn’t relative, as Tarski’s notion of sentential truth is, but it is restricted, since it’s only applicable to thoughts that are expressible in our own language. A true thought that isn’t so expressible will fail to satisfy the open sentence “(\Sigma p) ((x is the thought that p) \land p).” We surely don’t want to suppose that our language can express every thought, so we don’t want (\forall x) (x is a thought \to (\Sigma p) (x = \text{the thought that } p)). What substitutional quantification gives us is a partial definition, which says what it
is for a thought expressible in our own language to be true, but has nothing to say about the thoughts of dolphins:

\[(\forall x) \ (x \text{ is a thought expressible in our language} \rightarrow (x \text{ is true } \leftrightarrow (\Sigma p) ((x = \text{ the thought that } p) \land p))).\]

Since the expressions we can substitute for “\(p\)” are all meaningful sentences of our language, we have:

\[(\Pi p) (\exists x) \ (x \text{ is a thought expressible in our language} \land x = \text{the thought that } p)\]

Even within our own language, there’s a restriction that the sentence we substitute for “\(p\)” when we instantiate “\((\Pi p)____p____\)” can’t contain the substitutional quantifiers, lest we get circularities. In his earlier book, (2002, pp. 121ff), Hill presents an elegant proof that relaxing this requirement leads to antinomies. Thus our definition of truth needs a further restriction:

\[(\forall x) \ (x \text{ is a thought expressible in our language without using boldface variables} \rightarrow (x \text{ is true } \leftrightarrow (\Sigma p) ((x = \text{ the thought that } p) \land p))).\]

Quite apart from the paradoxes, there is an insidious circularity. We define truth using the substitutional quantifier but we only understand the substitutional quantifier because we’ve defined it in terms of truth. Hill’s response, following Gentzen’s (1969) lead, is to contend that the meaning of the substitutional quantifier is given by the natural-deduction rules of inference. Hill thinks of a natural-deduction proof as a sequence of thoughts and open thoughts. If, like me, you are puzzled by open thoughts—both about what they are and about what their inferential role might me—you can think of the open formulas that are written down in the course of a proof instrumentally, as uninterpreted expressions useful because they help us derive meaningful closed formulas that express valid propositions. Alternatively, we can follow Quine’s (1951, §15) example and formulate axioms of logic that don’t require free variables.

The natural-deduction rules for the ordinary predicate calculus don’t pin down a unique semantic interpretation for the quantifiers, because they don’t determine what the domain of discourse is to be. The domain has to include the objects named by the closed terms, but it might include many other things. There are similar limitations of the capability of the rules of inference to fix the meanings of the propositional quantifiers, Let \(L\) be our actual language and let \(M\) be an extension of \(L\), which expresses thoughts that \(L\) can’t express. \(M\) might be the language we’ll speak at some future time when our conceptual horizons have widened, or it might be a merely hypothetical expansion. Consider the following two candidates for the conditions under which \((\Pi p) \varphi (p)\) is true:

\[(\Pi p) \varphi (p) \text{ is true iff every sentence of } L \text{ obtained by substituting a sentence of } L \text{ for free occurrences of “} p\text{” in } \varphi (p) \text{ is true in } L.\]

\[(\Pi p) \varphi (p) \text{ is true iff every sentence of } M \text{ obtained by substituting a sentence of } M \text{ for free occurrences of “} p\text{” in } \varphi (p) \text{ is true in } M.\]
Both candidates legitimate precisely the same inferences in \( \mathcal{L} \). If all you know about \( \text{“}(\Pi p)\text{”} \) is the rules of inference, you’ll have no way of choosing between the two candidates. Yet the first candidate fits our intentions better than the second, since we want to say,

\[
(\Pi p) \ (\exists x) \ (x \text{ is a sentence of } \mathcal{L} \land x = \text{the thought that } p).
\]

There is no possibility of deducing this conclusion, using the laws of regular logic together with the rules for the substitutional quantifiers, from true premises about the “the thought that” function that are expressible without using the substitutional quantifiers; a compactness argument shows this. We can’t deduce the conclusion yet we can recognize it’s true, even though, supposedly, all we know about the substitutional quantifiers is given by the rules of deduction. I think we’re forced to the conclusion that our understanding of the substitutional quantifiers goes beyond what we are able to extract from the rules. The best we can hope for is that the rules get us started on the path toward an understanding of the substitutional quantifiers.

The indeterminacy of the range of the substitutional quantifiers is a problem for Hill, but it’s a problem faced by every inferentialist. Maybe it isn’t so bad. We can take substitutional quantification as primitive. We can learn to use the substitutional quantifiers the way we learn other parts of the language, without explicit instructions. Our starting point, in getting the hang of the correct usage of the substitutional quantifiers, is the rules of inference.

The remarkable thing about substitutional quantification is how little it demands, metaphysically. To get the substitutional version of quantification over propositions, we don’t need propositions, only sentences. To get the substitutional version of second-order logic, we don’t need classes or properties, only open sentences. That substitutional quantification is widely available, cost-free, is the conclusion we reach from Kripke’s (1976) negative answer to the question, “Is There a Problem About Substitutional Quantification?”

That substitutional quantification is metaphysically lighthanded makes little practical difference for Hill, who is fully committed to an ontology of propositions. It is useful, however, because it makes his results available to people like me, who are suspicious of the “third thing” ontology of propositions. There’s your thought that Cherry likes to chase squirrels and my thought that Cherry chase squirrels, but we doubt that there’s a third thing, the thought that Cherry likes to chase squirrels. The substitutional version of propositional quantification is available to us, nonetheless, since we harbor no doubts about the existence of sentences. Instead of \( \text{“}x = \text{the thought that } p\text{”} \), we’ll have a binary relation \( x \text{ is a thought that } p \).” With that relatively small change, we can accept Hill’s theory.

Going to the other extreme of metaphysical commitment, fully vested propositionalists can replace the substitutional quantifiers with objectual quantifiers ranging over all propositions, writing \( \text{“}(\forall p)\text{”} \) in place of \( \text{“}(\Pi p)\text{”} \) and \( \text{“}(\exists p)\text{”} \) in place of \( \text{“}(\Sigma p)\text{”} \). The resulting theory dances perilously close to the edge of paradox, but it’s not clear whether it falls in. Indeed, we have a consistency proof, of sorts. If we individuate propositions as coarsely as possible, declaring

\[
(\forall p) \ (\forall q) \ ((\text{the thought that } p = \text{the thought that } q) \leftrightarrow (p \leftrightarrow q)),
\]
we can eliminate the quantifiers, replacing \((\forall p) \varphi (p)\) with \((\varphi(0 = 0) \land \varphi(0 = 1))\), and replacing \((\exists p) \varphi (p)\) with \((\varphi(0 = 0) \lor \varphi(0 = 1))\). Of course, no propositionalist believes that there are only two propositions. Whether consistency can be maintained under a more believable way of individuating propositions is an interesting question, but one better pursued by somebody who doesn’t share my misgivings about the very existence of propositions.

5 Disquotational reference

Disquotational theories of truth have become quite popular. Partly as a result of their successes, disquotational accounts of reference have been less in fashion. The thought is that we supposed we needed the notions of reference and satisfaction to define truth, Tarski-style, but now that we have disquotational truth, we see that we don’t really need a notion of reference, disquotational or not. Hill stands out from the crowd as a strong advocate of disquotational reference.

We find an interesting example of how philosophers can draw opposite conclusions from a single body of evidence. How we refer to people, dogs, and planets is by means of a causal connection that links the referring name or concept to its bearer. For concrete things, names and/or concepts refer in virtue of an appropriate causal connection. There is no causal connection, or any relation remotely resembling a causal connection, joining numerals and numerical concepts to numbers. Consequently, Benacerraf (1965, 1973) concludes, numerical terms and concepts don’t refer. Reluctantly but irresistibly, he is driven to the conclusion that there are no numbers.

Hill (1987, p. 36) goes in the opposite direction. When we compare the relation the concepts we use to refer to people bear to the people they refer to and the relation numerical concepts bear to numbers, we find there is no comparison. The two relations are completely dissimilar. Of course, we can contrive a relation that connects the two kinds of reference, but it’s a contrivance. There is no remotely natural relation that subsumes both personal reference and numerical reference. The best response, Hill argues, is to stop looking for a natural relation. A disquotational conception gives us all the reference we need, without requiring us to seek a natural connection that isn’t there.

Substitutional quantification enables the disquotationalist to define the reference relation, and it’s determined by the definition that the concept of Barack Obama refers to Barack Obama, the concept of Peter Rabbit refers to Peter Rabbit, and the concept of the number 73 refers to the number 73.

On a disquotational conception of reference, we are assured that, provided Peter Rabbit exists, the concept of Peter Rabbit refers. Should we go on to say, as Hill (2014b, p. 70) does, that it refers determinately? It’s determined that there is a unique individual to whom the concept refers, but is there a unique individual of whom it’s determined that the concept refers to him? Compare: it is determined that one and only one individual will win the lottery versus there is one and only one individual of whom it is determined that she will win the lottery. We are assured of the truth of the judgment,

© Springer
The concept of Peter Rabbit refers to Peter Rabbit and to no one else, but that judgment only gives a uniquely determined referent to the concept of Peter Rabbit if we can be sure that, in the judgment, the term “Peter Rabbit” has a uniquely determined referent. Disquotationalism, as far as I can see, doesn’t entail a commitment one way or the other on the question whether there is something to which the concept of Peter Rabbit determinately refers.

The shocking possibility that, even though Peter Rabbit exists, the concept of Peter Rabbit might not have a determinate referent was raised by Quine’s (1960, ch. 2., and 1968) *gavagai* example. The example is flawed. Influential papers of Hill’s (1972) and Evans’ (1975) pointed out the formidable obstacles Quine faces if he tries to fill in the details of the example he sketched so faintly.

Permutation arguments, initiated by Richard Jeffrey (1964), seem to work better. Why did Mr. MacGregor pick up the watering can? The natural thing to say is that he thought that Peter was hiding underneath, but there’s something else we might say. Let’s say that \(x\) *u*\(^{-}\)hides under \(y\) if \(x\) and \(y\) are unit sets and \(x\)’s element hides under \(y\)’s element. We might say that Mr. MacGregor lifted the watering can thinking that Peter’s unit set was u-hiding under the can’s unit set. To report his mental state as a belief about Peter rather than about Peter’s unit set is more succinct and more natural, but if we want to be fully accurate, we want to know what about MacGregor’s mental state brings it about that he’s referring to Peter rather than his unit set, and we shouldn’t let ourselves be influenced by considerations of our convenience in reporting his mental states. Besides, if we spoke the u-language, in which the names we actually use to refer to people and rabbits were used instead to refer to their unit sets and in which simple predicates were used to ascribe properties and relations to unit sets of physical objects, then the succinct thing to say would be to report Mr. MacGregor as having a belief about Peter’s unit set.

My own inclination is to say that there’s no real difference between the two ways of reporting Mr. MacGregor’s beliefs. They are verbally distinct ways of saying the same thing. This, I take it, is Quine’s position, although he expresses it more provocatively.

There are two ways of describing MacGregor’s action. He wanted to catch Peter and thought Peter was hiding under the watering can. He wanted to u-catch Peter’s unit set and thought it was u-hiding under the watering can’s unit set. If what you want to know is whether Peter is caught and baked into a pie by Mrs. MacGregor, the way his father was, the distinction won’t make any difference. But it does make some philosophical difference. We like to think that our mental states, or rather our mental states embodied as they are in us in our environments, pick out a uniquely determined referent for each of our non-defective singular concepts. If the inscrutability of reference arguments are correct, we should abandon the “picking out” metaphor.

Quine reached the conclusion that reference is inscrutable on the basis of a premise of behaviorism, and many have taken its reliance on behaviorism to be the argument’s fatal flaw. Searle (1987) went so far as to declare the argument a *reductio* of its behaviorist premise. A closer inspection shows that the premise is inessential. What’s doing the work is a version of Frege’s (1884, p. x) context...
principle, that we should never seek the meaning of a word in isolation, but only in the context of a sentence.

The key observations are that words mean what they do because of the ways they are used, and that we use words by employing them in sentences. For Quine, “use” meant overt verbal behavior, but we don’t have to understand the notion so narrowly. We can count among the uses of a word its occurrence in a love sonnet you’re too shy ever to say out loud, or its use in thinking through a problem of interest only to yourself. Once it’s clear that, by the “use” of a word, we mean all the ways the word is used, that words mean what they do because of their use is seen as a harmless platitude.

We use words to perform speech acts, and we perform speech acts by uttering sentences. There are rare exceptions. We use the word “amo” as part of a paradigm that helps us remember how to conjugate verbs that end in “are.” But the exceptions are too few to make much difference.

If the inscrutability argument succeeds, then there are at least two choices for what “Peter Rabbit” refers to: Peter Rabbit and \{Peter Rabbit\}. Each choice performs equally well in accounting for the truth conditions, assertion conditions, and acceptance conditions for the sentences used to make assertions, and for the analogous conditions for the sentences used to perform the other speech acts. At the level of whole sentences, there is nothing to suggest that one reference theory gives a better semantic theory than the other. But nothing below the level of sentences is even relevant. Each candidate does equally well at accounting for the facts about use, and there are no facts about meaning that aren’t grounded in facts about use. Each candidate is just as good as the other. Moreover, our concept of Peter Rabbit is what we express when we use the name “Peter Rabbit,” so that, if it is undetermined what the name refers to, it is likewise undetermined what the concept refers to. The conclusion of the inscrutability argument is that neither the name nor the concept has a determinate referent.

Should we conclude that there’s no such thing as reference? That’s a reasonable conclusion, but not one that’s forced upon us. We are free to choose one of the candidates and to decide to use the word “refers” in such a way that it refers to the chosen candidates. Such a procedure would violate our expectation that the semantic attributes a linguistic theory ascribes to a language are posited because they are thought to reflect the practices of the speakers. We can avoid raising false expectations by explicitly declaring that attributions of reference are intended to have only instrumental value. Whereas that truth conditions proposed by the theory are intended to represent speaker usage, the reference conditions are recommended solely as a convenient means of getting to the truth conditions. With such a warning in place, there will be no harm in choosing and deploying a theory of reference.

If we decide to pick a candidate, which candidate should we choose? There is no possibility of choosing the candidate that best reflects the linguistic practices of the community; on that measure, the candidates score equally. The choice is guided solely by convenience, and the convenient choice, clearly, is to go along with folk usage and declare that “Peter Rabbit” refers to Peter Rabbit. In selecting the disquotational theory of reference over the various, equally empirically satisfactory, alternatives, we are, in Quine’s (1968, p. 49) phrase, “acquiescing in our mother tongue and taking its words at face value.”
For truth conditions, we had a choice: either to accept the disquotational theory as giving us everything we seek from a theory of truth, or to look deeper for a theory of truth conditions that gives causal explanations of verbal behavior. For the theory of reference, we have no such choice. A substantive, causally explanatory account isn’t available. Disquotational reference is the only game in town.

All this assumes that the inscrutability argument succeeds, which we can’t be sure of, since we can hardly claim to have examined every possibility for how usage and psychology might pin down a uniquely determined reference relation. Perhaps, if we dig deeper, we’ll find facts about thought and usage that determine answers to questions about what refers to what. A sensible methodology is to begin by defining the reference relation disquotationally, then dig beneath the surface to see if we can find a relation grounded in usage that underlies the relation we initially introduced disquotationally. Perhaps we’ll find one, perhaps we won’t, perhaps we’ll get a mixed outcome, for example, the conclusion that terms that purport to refer to physical objects have usage-determined referents, whereas mathematical terms do not. If we find a substantive, usage-based relation, it will explain the connections between language and thought and the world around us. Such a relation, if we find one, is sure to be more complicated and harder to grasp that the notion we get by simple disquotational postulation, but to abandon the quest for a substantive, causally explanatory understanding of the reference relation merely on the grounds that disquotational reference is simple and easy would be deplorably indolent.

The disquotational methods Hill develops are almost sure to be useful, both for understanding truth and for understanding reference. Disquotational semantics is a sensible starting point for our inquiries. With respect to truth, I have proposed that we ought to attempt to move beyond the starting point toward a causally explanatory understanding of truth conditions. We ought to do the same for reference too, if we can, but the inscrutability arguments give us reason to think that no more-than-deflationary understanding of reference is possible. Whether or not disquotational semantics is a satisfactory place to conclude our semantic inquiries, it is a sensible place to begin. We should, I think, welcome the techniques Hill has developed for disquotational semantics, whether or not we are willing to go along with the deflationary metaphysics.

Perhaps I might add, as a way of trying to deflate the deflationists’ pretensions, that, despite the deflationists’ claim to have lifted us above the shoals of metaphysics, deflationary metaphysics is still metaphysics.¹

References

Benacerraf, P. (1965). What numbers could not be. *Philosophical Review 74*: 47–73. (Reprinted in Benacerraf and Putnam [1985, pp. 272–294]).

¹ A version of this paper was presented at an APA symposium on Hill (2014a) in Vancouver in April 2015, where I received useful insights and advice from the other participants and from Professor Hill especially. I am grateful.
