Truth, Disquotation, and Expression
On McGinn's Theory of Truth*

Michael Glanzberg
MIT
glanzber@mit.edu

In *Logical Properties*, Colin McGinn offers a new theory of truth, which he describes as "thick disquotationalism." In keeping with wider theme of the book, truth emerges as conceptually primitive. Echoing Moore, it is simple and unanalyzable. Though truth cannot be analyzed, in the sense of giving a conceptual decomposition, McGinn argues that truth can be *defined*. A non-circular statement of its application conditions can be given. This makes truth a singularly remarkable property. Indeed, by McGinn's lights, it is the only property which enjoys being both unanalyzable and definable.[1]

McGinn argues that truth has a core feature which gives it this remarkable nature: the disquotation feature. This is the feature which allows us to step from $\ulcorner$It is true that $\phi\urcorner$ to $\ulcorner\phi\urcorner$. Truth is defined as the unique property for which this holds as an entailment. In McGinn's words, "truth is the (unique) property of a proposition from which one can deduce the fact stated by the proposition" (p. 97). Focusing on this sort of connection is also the hallmark of *deflationist* theories of truth. Strong forms of deflationism argue that truth is not a property at all, or not a 'substantial' property. McGinn argues that the central role of the disquotation feature does not imply any such deflationism. Truth is, according to McGinn a fully robust property, though one whose "essence" is disquotation (p. 87). (McGinn's disquotationalism is thus "thick," as opposed to deflationists who's disquotationalism is "thin.")

There is much in McGinn's basic outlook with which I am sympathetic. I agree with him that

*Thanks to Brad Armour-Garb, J. C. Beall, and Carol Voeller for helpful comments on earlier drafts of this paper.

[1] I shall not have space for a proper investigation of the idea of an unanalyzable but definable property. For an interesting discussion of Moore's version of the view that truth is simple and unanalyzable, see Cartwright (1987).

deflationism is too strong, and I also agree that the disquotational feature of truth is central to it. However, there are a number of points where I think we should ask about the details of McGinn's proposal. In the end, I shall suggest, the real work of enabling disquotation is done not by truth—the property of propositions *per se*—but by the semantics of a language used to express propositions. I shall also ask if we really can make do with just the step from ⌈It is true that ϕ⌉ to ⌈ϕ⌉. Typically, the converse is also taken as a characteristic mark of truth, though McGinn rejects it. I shall argue that on this point McGinn is mistaken, though once we see the relation between propositional truth and the rest of the semantics of a language, we will see more clearly why it is less problematic than McGinn holds.

I shall begin by considering McGinn's *uniqueness* claim: the claim that only truth has the disquotation feature. This is quite strong. McGinn considers some potential counter-examples, such as knowledge. We can step from ⌈It is known that ϕ⌉ to ⌈ϕ⌉. But McGinn suggests, this is simply because the factive 'known' decomposes into a combination of truth and other concepts, in such a way that truth alone is responsible for disquotation.

There are, however, a number of properties that are close to truth, and less clearly decomposable, which may raise questions about uniqueness. Consider, for instance, the concept of verification. Now, it may be that verification turns out to decompose as knowledge does. But at the very least, we do have some fairly clear models of verification-like concepts which do not have this feature. A straightforward one is that of proof in mathematics. For simplicity, let us take the concept of proof in some particular formal system. We may write this as ⌈$Prov(\ulcorner\phi\urcorner)$⌉. (We will always assume we are working over some appropriate base theory. PA would suffice.)

For the moment, it will be useful to compare provability with a sentential truth predicate ⌈$Tr(\ulcorner\phi\urcorner)$⌉. (Like McGinn, I ultimately prefer working with a truth predicate applied to propositions. I shall return to this in a moment.) Let us compare the properties of ⌈$Tr(\ulcorner\phi\urcorner)$⌉ and ⌈$Prov(\ulcorner\phi\urcorner)$⌉. It is commonly supposed that for truth, we will have the T-schema:

(T-S) $$Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi.$$

((T-S) for ‘T-Sentential’.) The disquotation feature is (in this setting) the *left-to-right* direction of (T-S). For reference’s sake, let us state this as a separate principle:

(D-S) $$Tr(\ulcorner\phi\urcorner) \rightarrow \phi.$$

For proof, the analog of (D-S) is a familiar principle:

(Rfn) $$Prov(\ulcorner\phi\urcorner) \rightarrow \phi$$

This is what is known as a *reflection principle*. It states the soundness—the correctness—of the theory for which $\lceil Prov \rceil$ is the provability predicate.

For a reasonable, sound base theory, $\lceil Prov \rceil$ is thus as much ‘disquotational’ as truth. Hence, we have a counter-example to McGinn’s uniqueness claim. But it appears we can still distinguish truth from proof (or more generally, truth from strong notions of verifiability). Roughly, proof is a narrower concept than truth. Those with any realist leanings at all will maintain that there are cases of true but unverifiable claims, and there are ready examples of true but unprovable sentences, as we learned from Gödel.

It appears to be the *right-to-left* direction of (T-S)—the one not included in the disquotation feature—which marks this difference. With concepts like provability and verifiability, we do not expect to have $\lceil \phi \rightarrow Prov(\ulcorner\phi\urcorner) \rceil$. This is a statement of completeness, and for any incomplete theory, it is just false. But we do expect to have $\lceil \phi \rightarrow Tr(\ulcorner\phi\urcorner) \rceil$, which reminds us that truth goes beyond provability.

I think what lies behind McGinn’s emphasis on the left-to-right half of the T-schema is in part the idea that this is the direction which really makes truth a useful concept. It is the aspect of disquotation we care about. Hence, it might be suggested that though we need to distinguish truth from proof or verifiability, that is not the most central aspect of truth. There is something right about this idea, but I still believe we cannot do without the right-to-left direction.

What is right about this idea is that in some crucial applications of the truth predicate, it is the left-to-right direction which does the real work. For instance, one of the usual applications of disquotation is to simulate infinite conjunctions. For instance, 'Everything John says is true' in effect expresses what is given by:

$$\bigwedge_{\phi} S(\ulcorner\phi\urcorner) \to \phi.$$

That 'Everything John says is true' implies this conjunction follows from each instance of $\ulcorner S(\ulcorner\phi\urcorner) \to Tr(\ulcorner\phi\urcorner)\urcorner$ and $\ulcorner Tr(\ulcorner\phi\urcorner) \to \phi\urcorner$, i.e. it only uses (D-S).[2]

But there is still much we would like to do with the truth predicate that is left out by (D-S). Simply on the basis of (D-S), we cannot come to *any* positive conclusion about truth at all. (Formally, it is easy to show that the base theory plus (D-S) fails to prove $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$ for any $\ulcorner\phi\urcorner$.) Ultimately, this will undermine the utility of truth as a device of infinite conjunction. In the example above, we start with $\ulcorner S(\ulcorner\phi\urcorner) \to Tr(\ulcorner\phi\urcorner)\urcorner$. But without more properties of truth than (D-S), we will never be able to reach any conclusions like this. For instance, we would be unable to gather evidence for them. Presumably we would come to $\ulcorner S(\ulcorner\phi\urcorner) \to Tr(\ulcorner\phi\urcorner)\urcorner$ by noting that the first thing John said was true, so was the second, etc. Eventually, we would use this to hazard the generalization that everything he says is true. But with only (D-S), we will not be able to reach even the initial observations that support the generalization. For, even if we recognize that $\ulcorner\phi\urcorner$ and $\ulcorner S(\ulcorner\phi\urcorner)\urcorner$ hold, we have no principle that licenses the conclusion $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$.

We thus need both directions of (T-S), to make use of the concept of truth, and to distinguish it from related concepts like proof. I shall suggest in a moment a way to understand the disquotation feature which makes room for both directions. But first, McGinn offers arguments against the right-to-left direction, and we should pause to consider them.[3]

McGinn offers two reasons to reject the right-to-left direction of (T-S). Both raise important points, but neither seems to me sufficient. Perhaps McGinn's primary concern is to reject what

[2] The exact sense in which truth predicates are devices of infinite conjunction is more subtle than I think is commonly recognized, but see Halbach (1999b).

[3] As J. C. Beall pointed out to me, we also need the right-to-left direction of (T-S) to understand such schematic generalizations as (Rfn). We interpret (Rfn) as saying that everything provable is true, which we get from the statement of (Rfn) together with the right-to-left direction of (T-S).

we might call the *equivalence* thesis. This is the thesis that goes along with the deflationist view that truth is not a property. It holds that the two sides of the T-schema express the same proposition, or are otherwise equivalent in a strong enough sense to indicate that the apparent mention of the property of truth on the left-hand-side does not really amount to a difference between the two sides. With McGinn, I am highly skeptical of this thesis. I believe McGinn is right that in virtue of predicating a property of a truth bearer, the left-hand-side has distinct content from the right-hand-side. But I do not see that this is a reason to reject either direction of (T-S). It only shows that we need to reject any bidirectional connection which is strong enough to amount to equivalence. (T-S) is put in terms of a material biconditional, and this is simply not strong enough to amount to the kind of equivalence we have rejected. There are even some modest strengthenings of the material biconditional which will still avoid equivalence. We can reject strong forms of deflationism and still accept the full T-schema in an appropriate form.

McGinn's second reason for rejecting the right-to-left direction of (T-S) is a direct argument that it does not generally hold. It fails, he argues, in problematic cases, such as truth-value gaps. This does raise an important issue, but not one, I suggest, which casts doubt on the principle of truth being bidirectional. Rather, the basic issue here is one of the conditional and related constructions in many-valued logics (or partial logics, or other frameworks suitable for truth-value gaps). McGinn notes that if $\ulcorner\phi\urcorner$ is assigned gap status, then the right-to-left conditional $\ulcorner\phi \rightarrow Tr(\ulcorner\phi\urcorner)\urcorner$ does not come out true. On perhaps the most common three-valued logic (the *strong Kleene* logic), this is correct (so long as $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$ cannot be assigned the value true when $\ulcorner\phi\urcorner$ is assigned gap). But the reason is that when we have assignments of gap, we have no tautologies at all. The propositional formula $\ulcorner p \rightarrow p\urcorner$ is not a tautology in the strong Kleene logic, for it comes out gap when $\ulcorner p\urcorner$ is assigned gap. We thus have every bit as much trouble with the left-to-right direction (D-S) McGinn does endorse. The natural assumption to make about the truth predicate in the presence of gaps is that it inherits its value (true, false, or gap) from the sentence it applies to. So if $\ulcorner\phi\urcorner$ is gap, so is $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$. But then $\ulcorner Tr(\ulcorner\phi\urcorner) \rightarrow \phi\urcorner$ is assigned gap by the strong Kleene logic just as the right-to-left direction is.[4]

Rather than showing that we need to give up on one direction of (T-S), the problem highlighted here is simply that of how to formulate (T-S) in a many-valued setting. There are a number of options. One is to leave (T-S) itself alone, and be more careful about its status. Though it is not true in any model, it never comes out false, and its two sides will always have the same value. Another option is to re-formulate it to directly express that we have entailments between its two sides. This can be done by re-casting (T-S) as a family of rules of inference which come out truth-preserving in many-valued or partial logics.[5]

If we admit truth-value gaps or other complications, the statement of the basic principles of truth likewise become more complicated. But this does not cast doubt on the basic idea of the bidirectional connection between $\ulcorner\phi\urcorner$ and $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$. In appropriate form, we still have (T-S) or something like it. And as we have seen, the right-to-left direction is important for distinguishing truth from other concepts like proof, or more generally verification, and for the utility of truth.

In arguing for both directions of the T-schema, I do not mean to suggest that disquotation is not an important aspect of truth. To see where it fits into the bidirectional picture, we should look more closely at how the T-schema works. So far, to ease the comparison with proof, I have taken the truth predicate to apply to sentences. McGinn takes it to apply to propositions, and we can gain some insight into disquotation by examining this version more closely.

For a sentence $\ulcorner\phi\urcorner$, the construction $\ulcorner$that $\phi\urcorner$ forms a name for the proposition expressed by $\ulcorner\phi\urcorner$. With this, a proposition version of the T-schema (basically McGinn's) is:

(T-P) $$Tr(\text{that } \ulcorner\phi\urcorner) \leftrightarrow \phi.$$

This schema, I shall argue, holds for two distinct reasons: one at the level of the propositions themselves, and one at the level of the connection between sentence and proposition. When

[4] There are some many-valued logics which avoid the lack of tautologies. Perhaps the most widely known is the Łukasiewicz logic, which sets $\ulcorner p \rightarrow p\urcorner$ to true when $\ulcorner p\urcorner$ is set to gap. But this undermines McGinn's argument against the right-to-left direction. Perhaps we could insist that $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$ is two-valued, but then it becomes unclear in what sense we really have truth-value gaps.

[5] For further discussion of these issues, see, for instance, Feferman (1984) and McGee (1991).

we look at propositions themselves, we find something related to (T-P), but which does not really have the importantly *disquotational* aspect. It is in the connection between sentence and proposition that we really find the basis of disquotation.

Suppose that propositions encode sets of truth conditions (at least). Then for our purposes, we may think of propositions as *functions* from truth conditions—possible worlds if we like—to two distinct values $\top$ and $\bot$. Let p be some proposition, where $p\colon W \to \{\top, \bot\}$. p is true if the actual circumstance, a distinguished truth condition a, is mapped to $\top$:

(P) $$Tr(p) \leftrightarrow p(a) = \top.$$

Of course, deciding whether p is true is then no easier than figuring out which circumstance amounts to a—deciding what is true requires a lot of knowledge about the world. But beyond that, the semantics of $\ulcorner Tr \urcorner$ is quite simple: evaluate a given function on a given input, and report the value.

The operation of evaluating p on the distinguished world a is familiar: it is the operation that returns the *extension* of p. In the notation of intensional logic, this is the operation ${}^{\vee}p$. When applied to a proposition, this operation returns a truth value, which can be put in a sentence position. We thus have:

$$Tr(p) \leftrightarrow {}^{\vee}p.$$

This may look like a form of the T-schema. It is the principle which fully describes the semantics of $\ulcorner Tr \urcorner$. But it is not a disquotation schema.

One way to see why not is to observe of how little use this schema is by itself. Even though it tells us exactly what the semantics of $\ulcorner Tr \urcorner$ is, it does not relate this semantics to anything we normally do in assertion. It only tells us how truth works with respect to propositions presented as functions on worlds. But we practically never encounter propositions presented directly in this way. Rather, we encounter them as the contents of *assertions*. We encounter the proposition *that* snow is white in the assertion of the sentence ‘Snow is white’, for instance.

Assertions *express* propositions via the sentences uttered. It is the relation of expression which gets us access to propositions in a useful way. Putting aside issues of context dependence, we may think of expression as a relation between a sentence and the proposition an assertion of that sentence expresses. If we are to make use of the truth predicate, we need to know how to apply it to propositions as they are expressed by sentences.

This is in fact what the propositional T-schema (T-P) does, though it does it in condensed form. (T-P) relies on the construction $\ulcorner$that $\phi\urcorner$. This is in essence a functor which takes us from a sentence, a linguistic item, to the proposition that sentence expresses. So (T-P) relates the proposition expressed to the sentence expressing it via the truth predicate. We can make this more explicit by making the contribution of 'that' more explicit. To do so, let us write the expression relation as $\ulcorner E(s,p)\urcorner$, which holds when sentence s expresses proposition p. Then we can make the content of (T-P) more explicit as:

(T-E) $$E(\ulcorner\phi\urcorner, p) \rightarrow (Tr(p) \leftrightarrow \phi).$$

The antecedent guarantees that the use of $\ulcorner\phi\urcorner$ on the right of the biconditional expresses exactly the proposition that appears in $\ulcorner Tr\urcorner$. With this and the fact that the truth value ($\top$ or $\bot$) of a sentence asserted is the value of the proposition it expresses, the biconditional on the right of (T-E) *follows* from (P).

(T-E) indeed has the kind of disquotational character McGinn highlights. $\ulcorner\phi\urcorner$ is *used* on the right, and thus talks about whatever it does—snow, grass, or sealing wax. It is connected by the biconditional to a claim entirely about a proposition. This allows for effortless movement between talk of a proposition and talk of the facts. But it is movement that is mediated by the *expression relation E*. It is because of the antecedent guarantee that the proposition we are talking about is the one expressed by the sentence that we can move so easily. The rest of the schema is provided by the simple property of truth of propositions (P); but the real work of making disquotation happen is done by $\ulcorner E\urcorner$.

The relation E is determined by the semantics of the language being spoken. Let us suppose, for argument's sake, that the semantics proceeds recursively (in a very roughly Tarskian way), by stating the referential properties of the lexical items, and then explaining how these combine to form meanings of sentences. (As we are talking about propositions, we should suppose that the semantics provides at least *intensions* for lexical items.) In this case, the relation *E* is determined by the relations between words and world(s) at the lexical level, and the ways these combine to determine what proposition is expressed by a complete sentence. These enable us to express propositions about the world by using sentences. The disquotational aspect of truth, I propose, primarily derives from this (with the help of (P)).[6]

With McGinn, I see little deflationary about this sort of view. In particular, we can now see that it is no more or less 'deflationary' than the semantics of the language is 'substantial'. I do think there is good reason to see the semantics of the language as substantial, rather than 'merely disquotational'. For one reason, the semantic properties of expressions are a contingent matter. If 'snow' had meant 'white' and 'grass' had meant 'green', for instance, 'Snow is white' would have expressed the proposition that grass is green.[7] At the very least, if we do have substantial referential properties determining the semantics of the lexical items, then there is little left of the idea that truth is 'deflationary'. In that case, even the disquotational aspects of (T-E) are based on highly non-deflationary semantic properties. As a number of authors have noted, the interesting part of a theory of truth is the theory of truth conditions—the theory of the relation *E*. Insofar as that can be appropriately deflationary, and support (T-E), we have a deflationary theory of truth. Insofar as it cannot (as I believe), we have a more substantial theory.[8]

I believe this puts the disquotational aspect of truth in a somewhat different light than McGinn's theory does. Insofar as the expression relation *E* is determined by the full semantics of the language, we have something stronger than the mere propositional or sentential T-

[6]Here I am in agreement with Higginbotham (1993).

[7]This is not quite as simple a point as it looks, for it is not entirely obvious what the bearers of this sort of modality are. A phonological shape can certainly change its semantic properties, but it is doubtful that a phonological shape by itself is close enough to our intuitive concept of 'word'.

[8]Hence, the most promising line for deflationists seems to me to reject the idea that semantics needs to be done in terms of reference, truth, and related notions. Presumably, such a view would also reject the sort of account of propositions I assumed above. (Field (1994) might be taken as a representative of this sort of approach to deflationism.) My primary claim here is that the interesting questions about truth, including disquotation and deflationism, come down to these sorts of questions about the semantics of a language.

I am unsure just where McGinn comes down on the nature of reference, but the discussion of predicates in Chapter 3 of *Logical Properties* certainly seems to be compatible with the more substantial, anti-deflationist approach.

schema. Informally, this is bought out by the way the semantics will lead us to claims $\ulcorner E(s, p) \urcorner$ via the internal structure of a sentence s. This has formal analogs as well. (Very roughly, whereas adding the sentential T-schema is a very weak extension of a theory (a conservative extension), adding the semantics which governs $\ulcorner E \urcorner$ is basically to add the clauses of a Tarskian truth definition. This is a much stronger theory. As Tarski himself observed, it is this theory which enables us to prove interesting generalizations about truth, such as that everything which is provable is true, or the law of bivalence.[9]) The approach to truth though the semantics of a language, to *expression*, and then to propositional truth preserves the disquotation feature, but as part of a much stronger theory.

I have argued that we need to have both directions of the T-schema to capture the notion of truth. I have further argued that the appropriate bidirectional version of the schema follows from the semantics of the language in question, together with some other basic principles. This certainly confirms McGinn's idea that truth supports disquotation, and that it is a key feature of truth that it does. But it seems to me that the real reason we have disquotation is not a singular and surprising property of truth, so much as a fundamental aspect of the semantics of the languages we speak.

References

Cartwright, R., 1987. A neglected theory of truth. In *Philosophical Essays*, pp. 71–93. Cambridge: MIT Press.

Feferman, S., 1984. Toward useful type-free theories, I. *Journal of Symbolic Logic* **49**:75–111.

Field, H., 1994. Deflationist views of meaning and content. *Mind* **103**:249–284.

Halbach, V., 1999a. Conservative theories of classical truth. *Studia Logica* **62**:353–370.

———, 1999b. Disquotationalism and infinite conjunctions. *Mind* **108**:1–22.

Higginbotham, J., 1993. Grammatical form and logical form. In J. E. Tomberlin, ed., *Philosophical Perspectives 7: Logic and Language*, pp. 173–196. Atascadero: Ridgeview.

McGee, V., 1991. *Truth, Vagueness, and Paradox*. Indianapolis: Hackett.

[9]For more details, see, for instance, Halbach (1999a).