



# Naturalism, Realism, and Normativity

**ABSTRACT:** *This essay describes three commitments that have become central to the author's philosophical outlook, namely, to liberal naturalism, to metaphysical realism, and to the epistemic and ontological objectivity of normative judgments. Liberal naturalism is contrasted with familiar scientific versions of naturalism and their project of forcing explanations in every field into models derived from one or another particular science. The form of metaphysical realism that the author endorses rejects every form of verificationism, including the author's one-time 'internal realism', and insists that our claims about the world are true or false and not just epistemically successful or unsuccessful and that the terms they contain typically refer to real entities. 'Representationalism is no sin'. The central part of the essay is an account of truth based on a detailed analysis of Tarski's theory of truth and of the insights we can get from it as well as of the respects in which Tarski is misleading. (This part goes beyond what the author has previously published on the subject.) The account of the objectivity of the normative in this essay draws on insights from Dewey as well as Scanlon.*

**KEYWORDS:** ethics, metaphysics, naturalism, philosophy of language, theory of truth

## 1. Naturalism

The terms 'naturalism' and 'realism' are useful (and even indispensable on occasion), but each of them has been applied to a wide range of different and even incompatible positions. I very much like the term 'liberal naturalism', which I first encountered in an important collection of essays edited by Mario De Caro and David Macarthur entitled *Naturalism in Question*. In my essay in that collection, Putnam (2004), I pointed out that while Boyd, Casper, and Trout's *The Philosophy of Science* (1991) admirably takes the trouble to tell us what the editors mean by 'naturalism' (and their naturalism is certainly not 'liberal'), their definition is disjunctive, and thus offers us two possible meanings of 'naturalism' rather than one. According to Boyd and colleagues, naturalism is '[t]he view that all phenomena are subject to natural laws, and/or [sic] that the methods of the natural sciences are applicable in every area of inquiry', and I argued that both disjuncts are extremely vague. In what sense does a 'naturalist' have to believe that the 'phenomenon' of Shakespeare's writing *Julius Caesar* was 'subject to natural laws'? Is it enough to believe that writing that play didn't *violate* any natural laws? Surely that makes it too easy to be a 'naturalist'! Or should one believe that in our 'first-grade'



conceptual system (Quine)—or in ‘the absolute conception of the world’ (Bernard Williams)—there is no fact of the matter as to what a single sentence of that play *means*, but the production of those marks on paper was a physical process whose physical explanation exhausts what there is to be known about the play? Or must one rather try to ‘naturalize’ questions about meaning, and so on? And what exactly does it mean to ‘apply’ the supposed ‘methods of the natural sciences’ to *Julius Caesar*? In fact, the sort of hard naturalism advocated by Boyd and Caspar’s volume chooses to say nothing about such subjects as literary criticism, and even the ‘Cornell Moral Realists’, a group of which Boyd is a leading member, regard talk of obligation (‘ought’) as belonging to the domain of the noncognitive. What makes them moral realists is the odd claim that ‘good’ refers to a natural kind, on all four feet with biological kinds. Instead of seizing either horn of this dilemma, De Caro and Macarthur pointed to and celebrated the existence of a number of philosophical positions that are ‘naturalist’ in the sense that John Dewey, the most famous naturalist philosopher of his period, was—namely, in the sense of rejecting all appeals to supernatural entities in philosophy while simultaneously rejecting the positivist demand that aesthetic and ethical concepts be reduced to the concepts of the natural sciences or expecting that they could, or eventually will be, so reduced. Nor do they accept the positivist view that history will eventually become a ‘science’. In their introduction to *Naturalism in Question*, De Caro and Macarthur emphasized that the liberal naturalism they advocate doesn’t regard normative utterances as somehow ‘second grade’ or merely ‘expressive’, but neither does it countenance a Platonic realm of normative facts independent of human practices and needs. At the same time, it does not countenance Moorean quasi-mystical faculties of moral intuition. All this I like very much.

I don’t know whether Tyler Burge would ever employ the term ‘liberal naturalism’, but I do know that his *Origins of Objectivity* (2010) is a great book. It takes indispensability for good scientific explanations (the sciences in question being vision science and related parts of psychology, in particular) to be sufficient reason for accepting an unreduced notion of representation in psychology and philosophy of mind, and I agree; at the same time, what Burge writes about ‘naturalization projects in philosophy’ is surely in the spirit of ‘liberal naturalism’:

Promoters of ‘naturalizing’ projects are often driven, I think, by misconceptions of science. These misconceptions breed misconceptions of mind. The notion of representation—of reference or attribution that can be correct or incorrect and that helps type-individuate kinds of psychological states—is entrenched not only in common-sense explanation but in scientific explanation in psychology. There is nothing unnatural or supernatural about such explanation. Some of the relevant psychology is well-supported, mathematically rigorous, mature science. There is no basis, even a *prima facie* one, to the worry that psychological notions are invitations to mystery or miracle. Even if there were such basis, the role that these notions play in powerful empirical science would undermine it . . . I know of no good ground

for thinking that... [psychologists'] explanatory claims must be twisted into the mold of biological or information-theoretic explanation, or any other explanation in the natural sciences, in order to be explanatorily successful. (Burge 2010: 296–7)

The reason I say that this is in the spirit of *liberal* naturalism is not just Burge's rejection of reductionism; some hard naturalists are quite proud of being 'nonreductive materialists', or the like. But even when hard naturalists eschew the demand that reference and representation talk be reduced to physical vocabulary, they feel the need to *explain* reference and representation in terms of 'causal relations' (see Boyd 2013: 51ff.). Even if the explanations are, as I believe, pure hand-waving, the spirit, if not the details, is obviously reductionist.

However, liberal naturalism, as represented by De Caro and Macarthur's two volumes (*Naturalism in Question* 2004 and *Naturalism and Normativity* 2010), covers a wide range of positions, including ones I disagree with. For example, one contributor to *Naturalism in Question*, Huw Price, rejects the whole idea that reference is a relation between linguistic items and worldly objects; an idea (he calls it 'representationalism') that is an essential part of my own realism; another contributor (and editor of the volume), advocates a version of Wittgensteinian 'quietism' (in Macarthur 2008) that I attack in Putnam (2008a). Thus, although I am indeed a 'liberal naturalist', that label does not say very much about my philosophy. To supplement it I need to say more, e.g., I am a liberal naturalist, but I am also a realist in my metaphysics and a realist with respect to the normative. This essay will attempt to sketch the outlines of a liberal naturalist metaphysically realist and normatively realist position in philosophy.

## 2. Metaphysically Realist

Readers who know of my philosophy from Wikipedia may be surprised to find me saying that my position is 'metaphysically realist'. Had I not consistently attacked 'metaphysical realism' from 'Realism and Reason' (1978) on? Well, not exactly. In that lecture and for a number of years afterward, I used 'metaphysical realism' (sometimes with capital letters) as a term for a specific position whose main feature was the insistence that the world can be divided into (mind-independent) objects and properties in exactly one way. But the use of that term was a mistake on my part. It was a mistake because, although I repeatedly explained what I meant by it, based on the natural understanding of the phrase 'metaphysical realism', it refers to a broad family of positions and not just to the *one* position I used it to refer to. In effect I *was* saying that by refuting the one philosophical view I called by that name I was ipso facto refuting anything that deserved to be called 'metaphysical realism', and that was not the case.

In addition, in Putnam (1994), my Dewey Lectures, which were the best known source for my post-'internal realist' views, I was still under the influence of Wittgensteinian quietism; although this is an aspect of Wittgenstein's thought I

grew increasingly unsympathetic to, it was not until my closing address, Putnam (2013), to my eightieth birthday conference in Dublin that I was willing to say, in those words, that I am a metaphysical realist without capital letters, a realist in my metaphysics, but not a ‘metaphysical realist’ in the technical sense I gave to that term in ‘Realism and Reason’ and subsequent publications. At the same time I was also willing to say, ‘Let me say for the record that I utterly and totally reject all versions of the ‘end of philosophy’ story, whether they come from Wittgensteinians, Rortians, Heideggerians, Derridians, or whoever. Philosophy was not a mistake; not in Parmenides’s time, not in Plato’s time, not in Aristotle’s or Descartes’s or Hume’s or Kant’s times, and not in our time. As Etienne Gilson put it, ‘philosophy always buries its undertakers’.

But what is it to be a metaphysical realist? In an elegant paper titled ‘Confessions of a Hard-Core, Unsophisticated Metaphysical Realist’, Maudlin (2015) describes his reaction to first reading the lecture in which I first used the term ‘internal realism’. In that lecture I had written:

The most important consequence of metaphysical realism is that *truth* is supposed to be *radically nonepistemic*—we might be ‘brains in a vat’ and so the theory that is ‘ideal’ from the point of view of operational utility, inner beauty and elegance, ‘plausibility’, simplicity, ‘conservatism’, etc. *might be false*. ‘Verified’ (in any operational sense) does not imply ‘true’, on the metaphysical realist picture, even in the ideal limit.

It is this feature that distinguishes metaphysical realism, as I am using the term, from the mere belief that there is an ideal theory (Peircean realism), or, more weakly, that an ideal theory is a regulative ideal presupposed by the notions ‘true’ and ‘objective’ as they have classically been understood. And it is this feature that I shall attack. (Putnam 1978: 125)

And Maudlin’s reaction:

We now have something quite concrete to discuss. The metaphysical realist thinks that an *operationally* ideal theory, a theory ideal *as far as we can tell*, might actually be false. This is a thesis that appears to be clear and concise. It is also a thesis that I take to be true and took to be true as a graduate student. So the most enduring impact ‘Realism and Reason’ had on me, despite its intent, was to make me a confirmed metaphysical realist. (Maudlin 2015: 485)

The thesis that a theory ‘ideal as far as we can tell’ might actually be false simply amounts to the claim that a statement that we are epistemically entitled to accept as part of our best theory of the world might be false, and there might be no way of verifying that it is false. And I agree with Maudlin that there are such statements.

Hence, if we understand metaphysical realism as he does, as simply making that claim, metaphysical realism is true. For example, there might be very strong reasons to believe that there are intelligent extraterrestrials, including radio signals or whatever, but the signals could be the product of quantum mechanical fluctuations of immensely low probability. And metaphysical realism in Maudlin's sense is part of my current view. Truth is not the same thing as warranted assertability under ideal conditions. 'Realism and Reason' was wrong.

But a responsible metaphysical realist needs to say something about what truth is and not simply what it isn't, and what follows is an attempt to do that, to say *something* (of course, not everything) about what truth is.

### 3. Truth: Tarski's Contribution and its Interpretation

In 'The Concept of Truth in Formalized Languages'<sup>1</sup> (hereinafter, Tarski 1983), Tarski immeasurably advanced our grasp of the formal logic of the predicate 'True sentence of *L*', where *L* is a formalized language, and by doing so he provided powerful new tools for mathematical logicians (something that is underappreciated by philosophers) as well as material for philosophers to assimilate and interpret. The tools for mathematical logicians that I refer to include techniques that were later adapted to formalize the statement that a set is definable over a given collection of sets with parameters from that collection and quantifiers ranging over that collection (tools that Gödel [1940] employed in the construction of the hierarchy of 'constructible sets' that he used to prove the consistency of the Axiom of Choice and the Generalized Continuum Hypothesis). These tools have been employed in many contexts ever since; the material for philosophers to assimilate and interpret includes the T-schema or truth schema in section 1 of Tarski (1983) (not to be confused with Convention T), Convention T, and the 'truth definitions', as they have come to be called, that are required to meet that criterion of adequacy. (In both the German and the original Polish of Tarski [1933], 'Convention T' is called a 'criterion' and not a 'convention'. 'Convention T' comes from the 1956 English translation, T being the first letter of the English word 'true' and W being the first letter of the German word '*wahr*'.) Although a minority of philosophers, including Popper, saw Tarski (1983) as putting forward a correspondence theory of truth, other philosophers have linked Tarski (1983) to 'deflationary' theories of truth, while the particular techniques Tarski employed led Hartry Field (1972) (correctly in my opinion) to see Tarski as finding an intimate link between *reference* (or, in Tarski's—and Field's—terminology, 'denotation') and *truth*; an insight that Field unfortunately linked to an unworkable 'naturalization project' (to use Burge's term), and later abandoned completely. The next two sections elaborate on these remarks.

<sup>1</sup> The original publication was in Polish, in 1933. The German translation appeared in 1935, and the 1956 English translation, titled 'The Concept of Truth in Formalized Languages', can be found in Tarski (1983), 152–278.

#### 4. Tarski's Criterion of Adequacy

Following Tarski, let us pretend that we have a particular interpreted formal language in mind, call it Bob. I choose an obvious proper name, Bob, and not  $L$ , because  $L$  looks like a *variable*—a variable for quantifying over formal languages—and it is essential to Tarski's approach that truth definitions are constructed *one language at a time*. Tarski (rightly) does not attempt to construct a definition of 'true in  $L$ ' when  $L$  is a variable over all languages, and doing so would immediately lead to paradoxes. We will assume that Bob has a finite number of primitive descriptive predicates and individual constants, etc. To avoid any appearance of circularity or begging the question, we will also assume that none of these is interpreted as referring to semantic properties or entities, such as truth or reference or 'correspondence' or facts or states of affairs. Let Meta-Bob be a second formal language that contains the whole vocabulary of Bob, one in which it is possible to quantify over inscriptions (sequences of symbols) in the language Bob. In addition, permit Meta-Bob to be set-theoretically more powerful than Bob in the ways needed for Tarski's purposes, but without having any primitive descriptive predicates or names other than those of Bob itself. A 'truth predicate' for Bob is a formula  $FORM(x)$  of Meta-Bob with one-free variable, say  $x$ . I avoid the custom of abbreviating the formula as 'True-in-Bob' (or even worse, as 'True-in- $L$ '), because doing so obscures the very issues I wish to raise. Of course, in ordinary language there is such a predicate as 'true in Bob', with no hyphens, but that predicate involves the ordinary language notion of truth, which Tarski claimed is inconsistent. That is the reason why we need a formal replacement for it. 'Defining truth in Bob', in Tarski's sense, means (1) finding a  $FORM(x)$  that is true of all and only those inscriptions  $x$  in Bob that are true in Bob and (2) reformulating this requirement in a way that does not involve using the everyday language word 'true' or any related notion. Tarski's solution to this latter desideratum, defining 'material adequacy' of a truth definition without using the word 'true', is the famous Convention T, which we state by saying:

(T) *FORM(x) is a materially adequate truth predicate for Bob just in case the result of replacing s by an arbitrary sentence of Bob and s\* by a structural-descriptive name of that sentence in:*

$$s \Leftrightarrow FORM(s^*)$$

is a theorem of Meta-Bob.

Tarski's notion of a 'structural-descriptive name' of an inscription is easiest to explain via an example: if  $s$  is 'Snow is white', then the formal name  $s^*$  is 'The letter capital  $S$  followed by the letter  $n$  followed by the letter  $o$  followed by the letter  $w$  followed by space followed by the letter  $i$  followed by the letter  $s$  followed by space followed by the letter  $w$  followed by the letter  $b$  followed by the letter  $i$  followed by the letter  $t$  followed by the letter  $e$ ' (see Tarski 1983: 157).

Thus, pretending that ‘Snow is white’ is a sentence of Bob:

*Convention T requires that* FORM( $x$ ) be defined in Meta-Bob in such a way that

‘Snow is white  $\Leftrightarrow$  FORM(The letter capital  $S$  followed by the letter  $n$  followed by the letter  $o$  followed by the letter  $w$  followed by space followed by the letter  $i$  followed by the letter  $s$  followed by space followed by the letter  $w$  followed by the letter  $h$  followed by the letter  $i$  followed by the letter  $t$  followed by the letter  $e$ )’ is *provable from Meta-Bob’s axioms*.

Readers who have gone through life believing that Convention T is a Davidsonian T-sentence or that it is an instance of ‘disquotation’ will be shocked to observe that ‘true’ does not occur in Convention T (although abbreviating Form( $x$ ) as ‘True-in-Bob( $x$ )’ can make this hard to see, which is why I didn’t do that)!

## 5. Field’s Reconstruction of Tarski: The Link between Reference and Truth

In his deservedly often-cited ‘Tarski’s Theory of Truth’ (Field 1972), Hartry Field argued that Tarski failed to reduce the notion of truth ‘to other nonsemantical notions’, but that what Tarski actually did was reduce the notion to other semantical notions.

This is an interesting reading, although I am not certain that Tarski had any such aim. What supports Field’s reading is that Tarski (1983) begins by describing the ordinary language notion of truth, and that this is where Tarski puts forward the T-schema (schema (2) on p. 155 of Tarski 1983). An instance of the latter is the famous ‘*Snow is white*’ is true iff *snow is white*; what perhaps counts against it is that at the end of that section (section 1) Tarski seems to give up the project of clarifying the ordinary language notion of truth entirely:

If these observations [concerning the inconsistency of the ordinary language notion—HP] are correct, the *very possibility of a consistent use of the expression ‘true sentence’ which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the very possibility of constructing a correct definition of this expression.* (Tarski 1983: 165, emphasis in original)

In any case, *one* of Tarski’s purposes, which he achieved, was certainly to find a way to define in Meta-Bob a predicate that *has the same extension* as the ordinary language semantical predicate ‘true in Bob’ (no matter which interpreted formalized language ‘Bob’ might be); as I put it in Putnam (1985), the predicate(s) FORM( $x$ ) we get by Tarski’s method do not express the same *concept* as the semantical predicate ‘true in Bob’. In Fregean language, FORM( $x$ ) has the right *Bedeutung*, but not the right *Sinn*. But by giving two of his papers the titles ‘The *Concept* of Truth in Formalized Languages’ and ‘The Semantical *Concept* of Truth’ [emphasis added], Tarski certainly gave the impression that his formal results somehow do

capture the *Sinn* of ‘true in Bob’ (that is, that they capture the *concept* of truth) and not just the extension of that everyday language semantical predicate.

## 6. A Technical Section

To explain all this, I need to say something about *how* one defines  $\text{FORM}(x)$  following Tarski’s recipe. Simplifying outrageously by pretending that all the predicates of Bob are monadic (the fact that this is in general not the case is the reason that all the mathematical ingenuity in Tarski’s paper was required), the easiest way to do this is to imagine that we construct Meta-Bob in two stages. I shall also, less outrageously, simplify by assuming Bob contains no names, only predicates, and (since the universal quantifier can be defined in terms of the existential quantifier  $[(x)F(x) \Leftrightarrow \neg(\text{Ex})\neg F(x)]$ ), I shall assume the primitive notation of Bob has only the existential quantifier.

*Stage 1.* First, (if necessary) we add to the vocabulary of Bob enough apparatus to do logical syntax (but not semantics!), that is, logical means for quantifying over strings of symbols, such as the letter capital  $S$  followed by the letter  $n$  followed by the letter  $o$  followed by the letter  $w$ , and over finite sequences of such strings. We also ensure it is possible to define such syntactic predicates as ‘ $x$  (a string of symbols) is a wff (well-formed formula) of Bob’ and ‘ $x$  (a string of wffs) is a proof in Bob’, ‘ $x$  is a theorem of Bob’, etc.

*Stage 2.* We temporarily add to Meta-Bob a primitive predicate  $\text{Ref}(x, y)$ , which we will read as ‘ $x$  (a wff of Bob with one free variable) refers to  $y$ ’. The interpretation of  $\text{Ref}$  should be clear from the following example: if  $x$  is ‘ $F(v)$ ’ and  $\underline{a}$  is an entity in the range of the quantifiers of Bob, then  $[\text{Ref}(x, \underline{a}) \Leftrightarrow F(\underline{a})]$ .

Next, we start to create an *inductive* definition of ‘ $\text{Ref}$ ’ by adding to the axioms (the axioms of Bob plus whatever axioms we need to do syntax and set theory in Meta-Bob) as follows:

For each *atomic* predicate (say, ‘ $\text{Glub}(x)$ ’) of Bob, we add as an axiom  $(y)[\text{Ref}(\text{‘Glub’ followed by ‘(followed by ‘}x\text{’ followed by ‘)’}, y) \Leftrightarrow F(y)]$ . In words, ‘ $\text{Glub}(x)$ ’ refers to  $y$  if and only if  $\text{Glub}(y)$ , and similarly for all the other atomic predicates of Bob. Note that there are only finitely many atomic predicates in what Tarski calls ‘formalized languages’, and hence only finitely many of these axioms. These axioms are the *basis clauses* of the inductive definition.

*Stage 3.* We add further axioms recursively extending the definition of ‘ $\text{Ref}$ ’ to molecular formulas, for example, an axiom saying that if  $x$  is the result of writing the disjunction symbol ‘ $\vee$ ’ between inscriptions  $w$  and  $z$ , then  $\text{Ref}(x, y) \Leftrightarrow \text{Ref}(x, w) \vee \text{Ref}(x, z)$ , and an axiom saying that if  $x$  is the result of writing the negation symbol ‘ $\neg$ ’ before  $w$ , then  $\text{Ref}(x, y) \Leftrightarrow \neg \text{Ref}(x, w)$ ,

AND (important!) we add to the axioms the statement, duly formalized, that for all  $w$ , if  $v$  is a variable and  $w$  is  $z$  preceded by an existential quantifier  $(\text{Ev})$  and  $z$  contains no free variables different from  $v$ , then  $\{\text{Ref}(w, \text{I}) \Leftrightarrow (\text{Ex}) \text{Ref}(z, x)\}$  &  $\{(\text{Ref}(w, \text{O}) \Leftrightarrow \neg(\text{Ex})(\text{Ref}(z, x)))\}$ . This amounts to adopting the convention that a *sentence* (a wff with no free variable) ‘refers to’  $\text{I}$  if it is true and to  $\text{O}$  if it isn’t.



Note that even if Bob's primitive predicates were all monadic, Bob would in general still contain polyadic defined predicates, for example, ' $F(x) \vee y=o$ ' and ' $F(x) \& F(y)$ '; therefore, we need a further clause in our inductive definition for existential quantification of a polyadic formula, but formulating such a clause requires more of Tarski's technique than we shall explain. Suffice it say that this technique involves constructing Ref as a relation between formulas and *sequences* (in fact, eventually constant infinite sequences) of objects.

*Stage 4.* Now comes the crucial stage: Given that 'Ref' has been inductively defined, it is a straightforward application of techniques familiar to logicians from Frege on to turn the inductive definition of 'Ref( $x,y$ )' into an *explicit* definition of a two-place predicate, say, SAT( $y,x$ ) [Read: 'y satisfies x', employing Tarski's terminology] provided Meta-Bob has strong enough set theoretic axioms; if it doesn't, we just add them. And *voilà*, we have a predicate of Meta-Bob whose definition involves only the atomic predicates of Bob and (if necessary) the primitive predicates of our favorite set theory *but no nonlogical predicates that weren't already in Bob*—a predicate that *provably* satisfies all the clauses in the inductive definition of Ref! N.B. that inductive definition depends only on the logical and syntactical axioms we added, but not on any 'semantical' axioms!

Therefore, we can drop Ref from our list of primitive predicates, drop the axioms that contained it, and use SAT( $y,x$ ) instead of Ref( $x,y$ ), since we know that all the old axioms for Ref are theorems of Meta-Bob, once we simply replace 'Ref( $x,y$ )' with SAT( $y,x$ ). And again—*voilà!*—it can be verified that if we take FORM( $x$ ) to be ' $x$  is a wff with no free variables & SAT( $\top,x$ )', then for each sentence of Bob, the result of substituting it for  $s$  and its structural-descriptive name for  $\hat{s}^*$  in 'FORM( $s^*$ )  $\Leftrightarrow s$ ' will be a theorem of Meta-Bob—Convention T is satisfied!

## 7. What Field Observed

What Field observed, and what I hoped to bring out clearly by breaking up the definition of SAT (Tarski's formal predicate for 'satisfaction') into stages in the way I did, was that ( $\top$ ) the 'truth predicate' (in the case of Bob, the predicate FORM( $x$ )) is defined in terms of the 'reference predicate' SAT, and (2) the construction can be said to *mimic* the process of defining truth in Bob in terms of reference in Bob (i.e., reference restricted to formulas of Bob); (3) reference in Bob (Field's 'denotation') is defined—or rather SAT, a relation *coextensive* with the converse of reference in Bob is defined—by turning an inductive definition into an explicit definition by well known means. Finally, (4) the basis clauses in the inductive definition 'define' reference in the case of *atomic* (undefined) predicates or rather specify the *extension* of reference in the case of such predicates by a finite list of cases. From this, Field concluded that Tarski reduced reference to what Field called 'primitive denotation' (denotation for atomic predicates) but that he failed to reduce 'primitive denotation' to nonsemantical notions (which, on Field's reading of Tarski 1983 was Tarski's aim), but only *listed* the extensions of the primitive predicates. Thus, according to Field, what remains to be done after Tarski is to 'naturalize' primitive denotation, that is, to define reference for atomic predicates in terms acceptable to

a materialist philosopher. A finite list of clauses such as ‘in Bob, “w” followed by “h” followed by “i” followed by “t” followed by “e” refers to all and only white objects’ is not a reduction at all, let alone one acceptable to a ‘physicalist’ like Field.

## 8. Philosophical Morals of the Above

(1) Apart from section 1, which ends on a massively pessimistic note, there is no attempt in Tarski (1983) to reduce any semantical notion to a nonsemantical notion (unless it be a failed attempt, which is what Field thought). What there is, is a definition of a predicate  $FORM(x)$  that is coextensive with ‘true in Bob’, but one that defines a property the sentence ‘snow is white’ has, for example, in a possible world in which snow is white and the word ‘white’ means black! (This is so because nothing in the definition of  $FORM(x)$  refers to the use or meaning of expressions.) As a *conceptual analysis of ‘true’*, the predicate  $FORM(x)$  fails miserably. (In 1953, Carnap suggested to me a way of meeting this objection. I describe Carnap’s objection [which depended on defining Bob by Bob’s ‘semantical rules’] and show why it fails in Putnam 1988: 61–67.) As I put it in Putnam (1985), people would not have had so much difficulty in grasping this undeniable fact if Tarski had titled his conception ‘The *Look No Semantics* Conception of Truth’. Of course, this is not to impugn the *mathematical* importance of the ‘material equivalence’—by which Tarski must mean the *coextensiveness*—of ‘true’ (in the relevant language Bob) and the truth predicate (our  $FORM(x)$ ) or the philosophical significance of Tarski’s work, which can only be extracted if we are clear on what it did and what it did not accomplish.

(2) ‘Correspondence theory of truth’ is the name traditionally given to theories of the form: ‘a statement [or thought or belief or proposition, etc., depending on the particular philosopher] is true if and only if the statement [or thought, etc.] corresponds to a fact [or to reality or to some appropriate piece of reality, depending again on the particular philosopher’s metaphysical views]’. The correspondence theory is discussed in section 1 of Tarski (1983) (where the T-schema [not Convention T!] is stated), but the section ends on a negative note about the whole project of clarifying the notion of truth in everyday language, as we noted above. However, Convention T itself does not (as I have emphasized) contain the word ‘true’.  $FORM(x)$  is a one-place predicate of Meta-Bob, and we have specified that Meta-Bob does not contain any words like ‘true’ nor any words like ‘fact’. Likewise, Convention T does not contain any words like ‘true’ nor any words like ‘fact’. Thus there is no way in which Meta-Bob, or Convention T, or the ‘truth predicate’  $FORM$  can even express the correspondence theory of truth.

(3) ‘Deflationary theory of truth’ is the name given to the more recent, twentieth-century theory that *the notion of truth is wholly captured* by one or the other of the two following disquotation principles:

(D) To call a statement (or sometimes a ‘proposition’, rarely a ‘sentence’ as in Tarski’s famous article) ‘true’ is simply to affirm the statement.

OR SOMETIMES,

(D') The statement '*S is true*'<sup>2</sup> is equivalent to the statement S.

Since Convention T does not *mention* truth, as I have been emphasizing, it obviously does not state either (D) or (D'). But it is plausible that it *presupposes* (D') for the following reason: Tarski's reader is supposed to see that if all of the conditionals that are required to be theorems of Meta-Bob by Convention T are theorems—and *if we assume that the theorems of Meta-Bob are true, which cannot be taken for granted* because a mathematical theory may contain false statements without being inconsistent—then each sentence of the form

$$\text{FORM}(s^*) \Leftrightarrow s$$

is *true*—and, using *the T-schema*

$$\text{true}(s^*) \Leftrightarrow s,$$

it follows that

$$\text{FORM}(s^*) \Leftrightarrow \text{true}(s^*).$$

Thus, it is *plausible* that the disquotation principle is presupposed by Tarski's claim that Convention T is a correct ('accurate', in the original Polish version) condition for the 'material adequacy' of a formula like  $\text{FORM}(x)$  as a truth predicate for Bob. In any case, the idea of disquotation easily arises from a study of Tarski's great paper. But there is all the difference in the world between accepting a disquotation principle and accepting the claim that such a principle captures completely what one has to know about truth. The latter is the thesis of deflationism. I conclude that Tarski is not committed to *that* thesis any more than he is committed to the correspondence theory of truth.

(4) Just as it is plausible to see a disquotation principle as presupposed by Convention T, even if Tarski did not state one, it is plausible to see the fact that the *extension of 'true in Bob' is determined by the extension of 'denotes in Bob'* as driving the entire strategy of defining the desired truth predicate ( $\text{FORM}(x)$ ) inductively in terms of a predicate (in our simplified version of Tarski, above,  $\text{SAT}(y,x)$ ) that is constructed to have precisely the extension of the everyday language predicate '*x refers to y in Bob*'.

In sum, and this is something I regard as of great importance, *Tarski's formal methods intuitively draw on and presuppose not just one property of truth, the T-schema, or disquotation, but on that property and the further property that the extension of 'true' depends on the extension of 'refers'. The concepts of truth and of reference are intimately related, and his entire procedure exploits the relation, as Field saw in 1972.*

<sup>2</sup> A technical problem: in (D') is 'S' a variable over statements? Or is (D') to be understood with some sort of systematic ambiguity? The literature discusses this problem extensively, and there are different proposals.

## 9. Realism is Incompatible with Deflationism

If one says ‘Asteroids [or daisies or marsupials] exist’, but one’s account of what it is to *understand* these speech sounds (or if one writes it instead of saying it out loud, if one’s account of what it is to understand such a string of symbols) does not mention any connection whatsoever between those ‘vocables’ (or those ‘strings’) and asteroids or daisies or marsupials or objects and properties in terms of which such entities can be described, then I, for one, fail to see how what one says can be understood in a realist way. This is a point I debated with Michael Devitt in Putnam (2013) as well as with Rorty in Putnam (2000), and I will not repeat all that here. What follows is addressed to a reader who ‘gets it’ and agrees. Of course, a *nonrealist* can be a deflationist and simply refuse to understand sentences about such things in a realist way. For a logical positivist to know the meaning of a sentence is just to know its method of verification, and the method of verification is to be described in terms of tests *we* can perform. For example, Carnap (1937: 37–38), discusses the statement ‘If all minds (or all living beings) should disappear from the universe, the stars would still go on in their courses’, and concludes that it is both cognitively meaningful and well confirmed.

Moreover, saying that certain sentences are *causally connected* to asteroids (or daisies or marsupials) isn’t enough to capture the way in which truth-apt assertions about real objects relate to the world. If I say that there are marsupials in Australia, I intend my utterance to be related to marsupials and not to anything else that the event of my making that utterance may have been caused by, for example, textbooks or zoos. In short, when I say it, I am *referring* to marsupials, and that fact is not captured by pointing out that my saying it was *causally connected* to marsupials.

If this is often underappreciated by deflationists, it seems to me that missing it comes from missing the point with which I closed the previous section and will repeat here:

Tarski’s formal methods intuitively draw on and presuppose not just one property of truth, the T-schema, or disquotation, but on that property **and** the further property that the extension of ‘true’ depends on the extension of ‘refers’ (and on the *possible* extensions of ‘refers’, if the logical vocabulary includes modal operators. Tarski did not consider such languages.) The concepts of truth and of reference are intimately interrelated.

If deflationists regularly fail to mention the interdependence of truth and reference, they do, however, recognize the need for some account of meaning or at least of the sameness of meaning. After all, I can speak of true sentences in a language that is not properly contained in my own language. The sentence “‘Schnee ist weiss’ is true if and only if Schnee ist weiss’ is not a well-formed sentence in either English or German. The standard form of disquotation in this case (a generalization of Tarski’s T-schema) is to say that if I am using English (or a formalized version thereof) as a metalanguage (for a part of German that is free of semantical words and includes the sentence ‘Schnee ist weiss’), then the appropriate T-sentence is:

‘Schnee ist weiss’ is true in German iff snow is white,

And, more generally, for any sentence *s* in the part of German in question,

(T) ‘*s*’ is true in German iff . . .

*where the three dots are to be replaced by the translation of the sentence *s* in English.*

That the notion of translation is needed for disquotation and therefore needed by deflationists (since their thesis is that grasp of disquotation is all that is needed for an understanding of truth) is widely recognized. But what I have not seen discussed by deflationists, let alone taken seriously, is the thought that *translating sentences presupposes knowing what their descriptive constituents refer to*. It is an illusion that disquotation does not presuppose the relation of reference.

The idea that one can speak of translation as if it didn’t involve reference can’t, of course, just stem from verificationism, as it did in Carnap, although Horwich, a leading deflationist, once suggested that sameness of meaning is just sameness of ‘use’ and coupled this with an explicitly verificationist account of ‘use’. Horwich (1990: 112) wrote, ‘The communal disposition to use a word in a certain way should not be regarded as simply the disposition to treat certain sentences as definitely and permanently acceptable and others not. In addition, there are dispositions to sanction various levels of confidence [*cached out as ‘betting behavior’*] in the truth of certain sentences—where the appropriate degrees of belief are a *function of observable circumstances*’ (emphasis added). Degrees of confirmation were also identified with betting behavior by Carnap.

Field (even after his conversion to deflationism) is no verificationist. But given that Field was a student of Quine’s (as well as of myself), I suspect that Quine’s account of ‘radical translation’ in *Word and Object* may well have influenced him, and in the next section I shall say why I think Quine’s account of radical translation was mistaken (something Burge does in detail in chapter 7 of *Origins of Objectivity*).

## 10. Quine’s Unrealistic Account of Linguistic Development and Translation

In Quine’s view, originally spelled out in *Word and Object* and developed and sometimes modified subsequently (in ways that do not affect the criticisms that follow), the translator who has no prior knowledge of the semantics of the language she is trying to figure out and children who are learning their first language have nothing to go on except sounds that their informants (parents, in the case of the child) are disposed to produce in response to stimulations of their nerve endings (their ‘speech dispositions’, a concept Quine takes from Skinner’s behaviorist psychology, as chapter 1 of *Word and Object* makes clear). To describe one of those dispositions as, in Quine’s famous example, ‘a disposition to say “gavagai”

when the speaker sees a rabbit', *as opposed to* 'a disposition to say "gavagai" when the speaker sees a collection of undetached rabbit parts' is regarded as an arbitrary imposition. The child, according to Quine (1958), does not discriminate rabbits as opposed to undetached rabbit parts, or 'rabbithood-exemplified', etc. in any determinate way, and the linguist's assumption that her 'jungle language' informant does so is likewise unwarranted.

It is easy to miss the fact that a very substantial position on the psychology of perception, child development, and language learning is presupposed by this Quinean account, one that supports Quine's whole indeterminacy thesis. As Burge (2010: 211–12) points out, in Quine's picture the sciences of linguistics and psychology float free of connections with biological accounts of the relations between the basic needs and activities of animals and the ecological facts about their environments. That picture also ignores psychological explanations of perception, which rely on biological accounts and feed back into them. No biological account describes the environment as containing undetached rabbit parts, and undetached rabbit parts do not figure in any scientific explanation in perceptual psychology either. When I say this, I am not turning my back on 'liberal naturalism' and opting for 'scientific naturalism'; I am rather opting for a *liberal naturalist understanding of what natural-scientific explanations are*, one that takes seriously what the natural sciences actually do, and does not require them to conform to 'physicalist' scruples (by avoiding notions like representation). Although Quine may have abandoned logical positivism, his insistence in *Word and Object* that the account of 'speech dispositions' begin with proximal stimuli ('nerve endings') and his view that children cannot be said to perceive objects until they have learned the apparatus of quantification are reminiscent of the positivist picture, in which a 'constructional system' is used to interpret bare sense data. There *is* a difference between linguistic reference, which presupposes complex cognitive capacities, and primitive perceptual representation, but the idea that organisms cannot discriminate objects prior to acquiring those cognitive and linguistic capacities is misguided. Linguistic reference grows out of perceptual representation, as Burge rightly argues, and if we can accept that, we need not be frightened into denying that truth of linguistic representations is a form of *accurate representation*. Moreover, if we see reference as growing out of perceptual transactions between organisms and things in their environments, we also won't be tempted to deny that reference *is* a relation between words and things, as Sellars at times did. Deflationism about reference typically goes with deflationism about truth. If saying that representation is a relation between organisms (and states of organisms and, derivatively, bits of language) and real things, properties, and events is 'representationalism', then representationalism is no sin!

## II. Normativity in Ethics

When philosophers teach ethics courses, they naturally emphasize the places at which the major theories and thinkers disagree. Less often do they try to see how insights from different schools of thought can be combined. Yet, it is especially important to do that in ethics, because when the subject matter is *how to live*,

as John Dewey emphasized in his sections of Dewey and Tufts' monumental *Ethics*, no one school has a monopoly on insight. Utilitarianism is inadequate as an account of our ethical lives, but it does identify a question that all too often needs be put to someone who proposes a significant rule of action, law, or policy: 'Would it contribute to the satisfaction of more people's desires or interests than the alternatives?' At the same time, Dewey (2009) emphasizes that desires and interests may be partial and selfish and that even when they are not, they may be based on bad reasoning or bad information. Desires and interests are not simply *givens* whose satisfaction needs to be maximized (as too many economists have assumed); they need to be subjected to intelligent and informed criticism. And Kantians have long pointed out that what satisfies the desires and interests of even a substantial majority may be wrong because it tramples on the rights of a minority. Although, in Putnam (2008b), I criticized Scanlon's (1998) claim that morality can be *founded* on the single principle that 'an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement', I do see that principle as a laudable attempt to state what is right in Kant's categorical imperative in a naturalistic spirit, that is, to state it in a way that does not depend on Kant's untenable belief that respect for the moral law has a 'supersensible' source (as contrasted to our 'inclinations' that have a natural source). Like Dewey ([1899] 1924: 283–84) I see the categorical imperative (as well as Scanlon's principle, just quoted) as 'a method for realizing the full meaning of a proposed course of action,' but not as the *foundation* on which morality rests. Morality rests on a complex and evolving set of human needs and interests, but those needs and interests are neither biological givens nor the product of mere contingency. As long as free inquiry and discussion survive, they will continue to be forced to meet the test of rational scrutiny, to answer the demand for good reasons.

What makes Scanlon's naturalism clearly a liberal naturalism is the willingness to use the notion of a *reason* without any accompanying 'naturalization project', and I am delighted that he chose to follow *What We Owe to Each Other* with a set of Locke Lectures entitled *Being Realistic About Reasons* devoted to the legitimacy and importance of doing just that.

## 12. In Closing

Part of what I have been saying might be summarized thus: the fact that something—perceptual representation or reference or truth or intentionality or reasons—can't be 'naturalized' in the way that 'physicalists' demand doesn't make those things 'non-natural' or 'queer' or suspiciously close to 'supernatural'. It is true that the notion of a reason, for example, is not the subject matter of a special science, but as I argued in Putnam (2002) that notion is presupposed by all science as well as by fields like history and politics and criticism (including philosophical criticism) that are not sciences, because in all of them one has to decide what there is *reason* to consider

and even what there is reason to test. (Notions like ‘plausibility’, ‘simplicity’, and ‘elegance’ figure in the *reasons* scientists give for testing certain theories at all. They are not scientific notions, but the activity of science presupposes a *reasonable* command of them.) Science depends on what is not fully scientific at every point. And both pragmatists and Wittgensteinians have rightly criticized the bad habit of turning what are sometimes useful distinctions into untenable dualisms, and that includes the dichotomy between normative and descriptive language (a dichotomy that our thick ethical concepts—including such ancient concepts as *brave* and *wise*—leap over without the slightest sign of embarrassment) and also includes the dichotomy between science and nonscience itself. One can learn from pragmatists and Wittgensteinians and philosophers of so many other kinds without becoming a card-carrying member of any philosophical sect. And that is something I have always tried to do in my philosophical life.

HILARY PUTNAM  
HARVARD UNIVERSITY  
[hilary.putnam@gmail.com](mailto:hilary.putnam@gmail.com)

## References

- Boyd, Richard, Philip Casper, and J. D. Trout, eds. (1991) *The Philosophy of Science*. Cambridge, MA: MIT Press.
- Boyd, Richard. (2013) ‘What of Pragmatism with the World Here?’. In Maria Baghramian (ed.), *Reading Putnam* (Abingdon: Routledge), 39–94.
- Burge, Tyler. (2010) *Origins of Objectivity*. Oxford: Clarendon Press.
- Carnap, Rudolf. (1937) ‘Testability and Meaning’ (part 2). *Philosophy of Science*, 3 (4), 420–71.
- De Caro, Mario, and David Macarthur. (2004) *Naturalism in Question*. Cambridge, MA: Harvard University Press.
- De Caro, Mario, and David Macarthur. (2010) *Naturalism and Normativity*. New York: Columbia University Press.
- Dewey, John. (1899–1924) *Ethics*. In Jo Ann Boydston (ed.), *The Middle Works, 1899–1924* (Carbondale, IL: Southern Illinois University Press), 283–84.
- Field, Hartry. (1972) ‘Tarski’s Theory of Truth’. *The Journal of Philosophy*, 69 (13), 347–75.
- Gödel, Kurt. (1940) *The Consistency of the Continuum Hypothesis*. Annals of Mathematics Studies 3. Princeton, NJ: Princeton University Press.
- Horwich, Paul. (1990) ‘Wittgenstein and Kripke on the Nature of Meaning’. *Mind and Language*, 2, 105–21.
- Macarthur, David. (2008) ‘Putnam, Pragmatism, and the Fate of Metaphysics’. *European Journal of Analytic Philosophy*, 4 (2), 33–46.
- Maudlin, Tim. (2015) ‘Confessions of a Hardcore, Unsophisticated Metaphysical Realist’. In *The Philosophy of Hilary Putnam*. Library of Living Philosophers. (Chicago, IL: Open Court), 482–96.
- Putnam, Hilary. (1994) ‘Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind’. *Journal of Philosophy*, 91 (9), 445–517. Reprinted in Hilary Putnam, *The Threefold Cord* (New York: Columbia University Press, 1999), 3–70.
- Putnam, Hilary. (1978) ‘Realism and Reason’. In *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul), 123–38.
- Putnam, Hilary. (1985) ‘A Comparison of Something with Something Else’. *New Literary History*, 17 (1), 61–79. Reprinted in Hilary Putnam, *Words and Life* (Cambridge, MA: Harvard University Press, 1994), 330–50.



- Putnam, Hilary. (1988) *Representation and Reality*. Cambridge, MA: MIT Press.
- Putnam, Hilary. (2000) 'Richard Rorty on Reality and Justification'. In Robert Brandom (ed.), *Rorty and his Critics* (Oxford: Blackwell), 81–87.
- Putnam, Hilary. (2002) 'The Philosophers of Science's Evasion of Values'. In *The Collapse of the Fact-Value Dichotomy* (Cambridge, MA: Harvard University Press), 135–45.
- Putnam, Hilary. (2008a) 'Reply to David Macarthur'. *European Journal of Analytic Philosophy*, 4 (2), 47–49.
- Putnam, Hilary. (2008b) 'Capabilities and Two Ethical Theories'. *Journal of Human Development*, 9 (3), 377–88. Reprinted in Hilary Putnam, *Philosophy in an Age of Science*. Edited by Mario De Caro and David Macarthur (Cambridge, MA: Harvard University Press, 2012), 299–311.
- Putnam, Hilary. (2009) 'Dewey's Central Insight'. In Larry A. Hickman and Giuseppe Spadafora (eds.), *John Dewey's Educational Philosophy in International Perspective: A New Democracy for the Twenty-First Century* (Carbondale: Southern Illinois University Press, 2009), 7–21.
- Putnam, Hilary. (2013) 'From Quantum Mechanics to Ethics and Back Again'. In Maria Baghramian (ed.), *Reading Putnam* (Abingdon: Routledge), 19–36.
- Putnam, Hilary. (2013) 'Comment on Michael Devitt'. In Maria Baghramian (ed.), *Reading Putnam* (Abingdon: Routledge), 121–26.
- Quine, Willard Van Orman. (1958) 'Speaking of Objects'. *Proceedings and Addresses of the American Philosophical Association*, 31, 5–22. Reprinted in Willard Van Orman Quine, *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 1–25.
- Scanlon, Thomas. (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Tarski, Alfred. (1983) *Logic, Semantics, Metamathematics*. Edited by John Corcoran. 2d ed. Indianapolis: Hackett.