

# 9

## Denotation and Discovery

Gary Ebbs

“Denotation and Discovery,” by Gary Ebbs, published by Rowman & Littlefield Publishers, Inc., appears here by permission of the author and publisher.

This material is protected by copyright. All rights reserved. Please contact the publisher for permission to copy, distribute or reprint.

### PRACTICAL JUDGMENTS OF SAMENESS OF DENOTATION

Speakers of the same natural language typically take each other's words at face value. If you show me a ring and say, “This ring is gold,” I'll take you to have asserted that the ring is gold, without first thinking about whether this way of taking your words is justified. When I take your word ‘gold’ at face value in this way, I take it to be true of the ring just in case *my* word ‘gold’ is true of the ring, hence just in case the ring is gold. This way of taking your word is *like* a judgment in that I may revise it in light of new information, but *unlike* a judgment in that it is unreflective and may never come up for review. I call it a *practical judgment of sameness of denotation*.

Our practice of taking other English speaker's words at face value extends across time, from moment to moment and, in some cases, for centuries. For example, if we learn that in 1650 a jeweler showed John Locke a ring and said, “This ring is gold,” we'll take the jeweler to have asserted that the ring is gold. When we take the jeweler's word ‘gold’ at face value in this way, we in effect take it to be true of an object *x* just in case *our* word ‘gold’ is true of *x*, hence just in case *x* is gold. We thereby make what I call a *practical judgment of sameness of denotation across time*.<sup>1</sup>

We express such judgments when we take ourselves to have made a discovery or to agree or disagree with other speakers. For instance, we take for granted that when chemists first accepted the sentence “Gold is the element with atomic number 79,” they did not thereby introduce a new denotation for the word ‘gold’.<sup>2</sup> Trusting our practical judgments of sameness of denotation for ‘gold’, and taking for granted that gold is the element with atomic number 79, we can *agree* with a jeweler who in 1650 showed John Locke a

ring and said, "This ring is gold," only if the ring was (a bit of) the element with atomic number 79. We realize that perhaps the ring wasn't gold, even if it passed all of the jeweler's tests for being gold. In this practical way we acknowledge that truth is independent of belief.

There is a diverse group of philosophers, including David Chalmers, Michael Dummett, Frank Jackson, John McDowell, Christopher Peacocke, and (sometimes) Hilary Putnam, who think that if no theory of what determines the denotations of our words *justifies* (or shows that we are *entitled* to) our practical judgments of sameness of denotation, then our *impressions* that we can make discoveries, that we can agree or disagree, and that truth is independent of belief, are *illusory*.<sup>3</sup> Against this, I will emphasize that we trust our practical judgments of sameness of denotation more than any theory of what determines the denotations of our words. I will argue that the denotations of our words are not settled by *a priori* linguistic rules or by causal-historical facts, but that our practical judgments of sameness of denotation are of a piece with our pursuit of truth, so we should not reject them. Instead, we should build these judgments into a deflationary account of denotation and truth.

### ANALYTICITY

To begin with, consider the relationship between our practical judgments of sameness of denotation and the thesis that some of our sentences are "analytic," in the sense that we cannot abandon them without changing the subject. One version of this thesis is that we *make* some of our sentences true by agreeing on how they are to be evaluated.<sup>4</sup> The main problem with this version of the thesis is that it conflates truth and belief. As Frege observed in a different connection, "Being true is different from being taken to be true, whether by one or many or everybody, and in no case is to be reduced to it. There is no contradiction in something's being true which everybody takes to be false" (1964, p. 13).

One way to support the thesis that some of our sentences are analytic without conflating truth and belief is to derive the thesis from a description of how we evaluate sentences. To see how this might be done, consider W. V. Quine's account of the deviant logician's predicament. Against the idea that deviant logicians may "reject the law of non-contradiction and accept an occasional sentence and its negation both as true," Quine argues as follows:

[They] think they're talking about negation, '¬', 'not'; but surely the notation ceased to be recognizable as negation when they took to regarding some conjunctions of the form 'p·¬p' as true, and stopped regarding such sentences as implying all others. Here, evidently, is the deviant logician's predicament: when he tries to deny the doctrine he only changes the subject. (1986, p. 81)

The moral is that even though truth is not up to us, for some words, including '¬', 'not', we can agree on criteria that settle whether or not a speaker is using them to talk about the same subjects that we talk about when we use them.

This is the truism behind what I call methodological analyticity—the idea that even though truth is not up to us, there are sentences we cannot reject without changing the subject. The least problematic version of the idea, due to Rudolf Carnap, makes sense only for sentences of an artificial language system.<sup>5</sup> If we accept Quine's textbook explanations of '¬', sentences of the form '¬(p·¬p)' come close to being "analytic" in Carnap's strict sense of that troublesome word.

The sort of methodological analyticity I want to discuss is also supposed to be a feature of natural language sentences, such as 'Bachelors are unmarried adult males'. The idea is that we *tacitly* agree on criteria that settle whether or not a speaker is using 'bachelor' to talk about bachelors, 'adult' to talk about adults, and so on. We tacitly agree, for instance, that no one can reject 'Bachelors are unmarried adult males' without changing the subject. Moreover, the criteria on which we tacitly agree are in principle obvious to us—we can tell by reflecting on our own usage of the terms whether or not an explicit statement of the criteria is correct. Natural language sentences that are analytic in the sense can play a methodological role in our inquiries that is analogous to the more strictly defined methodological role of analytic sentences in Carnap's artificial constructed language systems. That is why I call this sort of analyticity *methodological*.

One might be inclined to dismiss methodological analyticity (even the pure form of it that is restricted to artificial languages) with the claim that for any sentence we accept, we can imagine that it's false. This claim may seem to follow immediately from Frege's distinction between truth and belief. But that distinction has no direct bearing on whether we can imagine that a sentence we now accept is false. What would it be, for instance, to imagine that a sentence of the form '¬(p·¬p)' is false?

A better argument against methodological analyticity is that we may at one time feel confident that we could not reject a given statement without changing the subject, but later realize that we were wrong. This point is usually attributed to Quine, but Hilary Putnam was the first to present examples that make it convincing.<sup>6</sup> He observed, for instance, that in the eighteenth century, scientists had no idea how their theory that physical space is Euclidean could be false—no idea how one could reject the statement that physical space is Euclidean without changing the subject. After much theoretical work in mathematics and physics, scientists replaced that earlier theory of space with the theory that physical space is non-Euclidean. When scientists came to believe that physical space is non-Euclidean, they took themselves not to have *changed the topic*, but to have *discovered* that space is non-Euclidean.

To accept this description of the case, we must trust the later scientists' practical judgments of sameness of denotation for the phrase 'physical space' more than we trust the earlier scientists' speculation that one cannot reject the statement that physical space is Euclidean without changing the subject. The later scientists might be wrong about physical space—perhaps it's Euclidean after all. But we take them to have *discovered* that physical space is non-Euclidean, and so we accept their practical judgments of sameness of denotation for 'physical space'. These practical judgments are embodied in their use of that phrase to express what they take to be a discovery about physical space. Their use of that phrase links it to earlier uses of the phrase, and those earlier uses of the phrase are linked to even earlier uses of it. Taken together, these uses of 'physical space' constitute a transtemporal chain of practical judgments of sameness of denotation for 'physical space'. In a similar way, every inquiry brings with it some chain or other of practical judgments of sameness of denotation across time. In this sense, our practical judgments of sameness of denotation are of a piece with our pursuit of truth.

One might think that this reasoning only shows that we can be radically wrong about our own tacit criteria for applying our terms, not that methodological analyticity is incorrect. As I defined it above, however, methodological analyticity implies that we can tell just by reflecting on our own current usage of a term whether or not a given explicit statement of how it should be applied is correct. Putnam's counterexamples show that we can't tell just by reflecting on our own current usage of a term whether or not a given explicit statement about how it should be applied will survive a conflict with our practical judgments of sameness of denotation. In this way, Putnam's counterexamples undermine methodological analyticity.

### PRIMARY INTENSIONS

One might grant that we *trust* our practical judgments of sameness of denotation, but think that if we cannot also *justify* them, then our impression that we can make discoveries is illusory. It is natural to think that for every word used by a given speaker, there are linguistic rules that determine the denotation of that word. If there are such rules, then, for instance, my practical judgment that another speaker's word 'gold' has the same denotation as my word 'gold' is correct if and only if the denotation determined by the linguistic rules for her word 'gold' is the same as the denotation determined by the linguistic rules for my word 'gold'. Can we identify linguistic rules that we can use to justify our practical judgments of sameness of denotation?

We cannot use disquotational rules to justify these judgments. A disquotational rule for applying a given word—a rule such as 'gold' denotes an ob-

ject  $x$  if and only if  $x$  is gold—tells us nothing about the conditions under which a practical judgment of sameness of denotation for that word is true. A justification of such judgments therefore requires *more* than a disquotational specification of rules for applying our words.

Assuming that for every word  $w$  used by a given speaker  $S$  there are rules that determine the denotation of  $w$ , it is tempting to think that those rules are settled by  $S$ 's *beliefs* about how  $w$  should be applied. Inspired by this thought, David Chalmers has recently proposed a theory of intensions (or concepts) that is meant in part to justify our practical judgments of sameness of denotation across time. He defines the *primary intension* of a word as a special sort of function from (agent-centered) worlds to extensions: in a given (agent-centered) world, the primary intension of a word picks out what the extension of the word would be if that (agent-centered) world turned out to be actual (Chalmers 1996, p. 57). To grasp the primary intension of 'water', for instance, we must grasp a function that yields the set of all and only portions of *water* as value if the actual (agent-centered) world has water in its rivers, lakes, and oceans, but yields the set of all and only portions of *twin-water* as value if the actual (agent-centered) world has twin-water in its rivers, lakes, and oceans.

What is distinctive of a primary intension, according to Chalmers, is that our grasp of it is independent of *all* our empirical beliefs. He argues that there *must be* a primary intension for any word that we can use to express a discovery. If we are to express a discovery about water that is based on our examination of a given sample of what we take to be water, he reasons, we must be able to say *why* it counts as a sample of *water* by appealing to rules that we can grasp without going through any empirical investigation or presupposing any empirical beliefs (p. 62).

This reasoning moves from a truism to a substantive epistemological claim. The truism is that we take ourselves to express a discovery about water by using the term 'water' only if we take for granted that the denotation of 'water' does not change as a result of our supposed discovery. The substantive epistemological claim is that we are entitled to take ourselves to express a discovery about water by using the term 'water' only if we can *justify* our practical judgments of sameness of denotation for 'water' by appealing to rules that we can grasp without going through any empirical investigation or presupposing any empirical beliefs.<sup>7</sup>

A primary intension is well-suited to this justificatory role, according to Chalmers. "The intension specifies how reference depends on the way the external world turns out, so does not itself depend on the way the external world turns out" (p. 57). By reasoning about "what our words *would* refer to if the actual world turned out in various ways," Chalmers thinks, we can simultaneously see that our words have primary intensions and *discover* what they are.

The main problem with this proposal is that what we *actually* say when we find ourselves in a previously imagined situation almost always trumps our earlier speculations about what we *would* say if we *were* to find ourselves in that situation. What we actually say when we find ourselves in a previously imagined situation reflects our best current judgment of what is true in that situation. When we are actually in the previously imagined situation, our best judgment of what is true in that situation brings with it practical judgments of sameness of denotation. If those practical judgments of sameness of denotation conflict with earlier speculations, then so much the worse for those speculations. A scientist in the eighteenth century might have confidently predicted that even if it turns out that there is some mathematically consistent non-Euclidean geometry, it *can't be* the description of what he calls physical space. But the actual history of our practical judgments of sameness of denotation for 'physical space' yields the opposite conclusion that when scientists came to accept that physical space is non-Euclidean, they did not thereby change the subject by tacitly defining a new denotation for the phrase 'physical space'. As I emphasized above, the later scientists' practical judgments of sameness of denotation are of a piece with their inquiry into the shape of physical space. We take the later scientists' practical judgments of sameness of denotation to trump the earlier scientists' speculations because we are confident that the later scientists have *discovered* that physical space is non-Euclidean.

This example shows that statements we can't imagine giving up without changing the subject are not thereby *guaranteed* to be true. But one might think that to accept Chalmers's claim that some of our words have primary intensions that we can know *a priori*, we need only suppose that some of the statements that we can't imagine giving up without changing the subject actually *are* true. The trouble is that Chalmers's primary intensions are supposed to "back *a priori* truths"—statements that are "true no matter how the actual world turns out" (p. 59). Hence to accept Chalmers's claim that some of our words have primary intensions that we can know *a priori*, it is not enough to suppose that some of the statements that we can't imagine giving up without changing the subject are true.<sup>8</sup>

Like Chalmers, Frank Jackson tries to defend the inference from "we don't understand how we could give up statement S without changing the subject" to "we could not give up statement S without changing the subject." "[S]urely it is possible to change the subject," Jackson reasons, "and how else could one do it other than by abandoning what is most central to defining one's subject? Would a better way of changing the subject be to abandon what is less central?" (Jackson 1998, p. 38). The mistake here is to suppose that our best *current* judgment about what counts as changing the subject is immune to future revisions. It is a truism that if we want to change the subject, we must rely on our understanding of what is most

central to defining it. But this truism does not establish that our current understanding of what is most central to defining our subject cannot be revised without changing the subject. This claim is discredited by many actual cases in which we were once confident that we could not revise a given statement without changing the subject, but discovered later that we were wrong.

Both Jackson and Chalmers sometimes claim to be able to accommodate such discoveries (Chalmers 1996, pp. 55–6; Jackson 1998, pp. 46–55).<sup>9</sup> They do not seem to realize that to accommodate such discoveries is to concede that our practical judgments of sameness of denotations are more trustworthy than our explications of our own primary intensions. This concession undermines their thesis that our understanding of our own primary intensions justifies our practical judgments of sameness of denotation across time.<sup>10</sup>

### THE CAUSAL-HISTORICAL THEORY OF DENOTATION

You may be convinced by these arguments against Chalmers and Jackson but hope to find a different sort of justification of our practical judgments of sameness of denotation. Saul Kripke and Hilary Putnam devised the causal-historical theory of denotation to explain and justify the practical judgments of sameness of denotation that led them to reject the description theory of proper names and natural kind terms. In outline, the causal-historical theory is that the denotation of a name or a kind term is *initially* determined by an "ostensive definition" that may partly rely on causal connections to samples or things that the name or term is to denote. Once the denotation of a word is established in this way, speakers of the language can be credited with using that word if and only if they are connected by an "appropriate" causal chain with other speakers who use the word, and they are "minimally competent" in its use. "Appropriate" and "minimally competent" are place-holders for specifications of the causal chains and competencies that explain and support our practical judgments of sameness of denotation.

There is no consensus about how to specify these supposed chains and competencies, and many philosophers concede that there are kind terms (such as names of biological species) that pose challenges for the causal-historical theory. Nevertheless, many philosophers still think that by citing causal-historical facts we can explain why the denotations of our word 'gold', for instance, did not change as a result of our discovery that gold is the element with atomic number 79. Even in this paradigm case, however, as I shall now try to show, we cannot justify our practical judgments of sameness of denotation by citing causal-historical facts.

### A THOUGHT EXPERIMENT<sup>11</sup>

The historical background for my argument is that platinum was not discovered until the mid-eighteenth century, when chemists called it "white gold" because of its striking similarities to what they previously called gold.<sup>12</sup> Platinum has a higher melting point than gold. But like gold, platinum dissolves in aqua regia, which was named for its ability to dissolve gold.<sup>13</sup> In 1650, a chemist applying this "acid test" to a sample of platinum might have concluded that it should be called gold.<sup>14</sup> We now know that platinum and gold are different elements: platinum is the element with atomic number 78, and gold is the element with atomic number 79.

With this in mind, suppose that there is a Twin Earth that is indistinguishable from Earth up until 1651, when large deposits of platinum are uncovered in Twin South Africa, and that once it is established by Twin Earth chemists that the newly uncovered metal dissolves in aqua regia, members of the Twin English-speaking community call it 'gold,' treating it in the same way we treat gold: the platinum is mined as gold, hammered (and later melted) together with gold to produce coins and bars that are valued by Twin Earthlings just as we value gold. Everyone on Twin Earth trusts the Twin Earth chemists' judgment that the newly uncovered metal is properly called 'gold.'

Suppose also that on Twin Earth chemistry develops in almost exactly the same way in which it develops on Earth, except that when Twin Earth chemists investigate what they call 'gold', they conclude that there are two kinds of 'gold'—their word 'gold' denotes *x* if and only if *x* is (a bit of) the element with atomic number 78 or *x* is (a bit of) the element with atomic number 79.

Recall that Twin Earth is just like Earth with a slightly different future after platinum is first uncovered in Twin South Africa in 1651. To see the possibility of this Twin Earth scenario, it is enough to imagine a few accidental differences between the two communities that allow for the uncovering of large amounts of platinum on Twin Earth.<sup>15</sup>

The crucial point is that just as members of our English-speaking community take for granted that the denotation of the English word 'gold' did not change as a result of the discovery that it denotes *x* if and only if *x* is (a bit of) the element with atomic number 79, so members of the Twin English-speaking linguistic community take for granted that the denotation of their Twin English word 'gold' did not change as a result of their discovery that it denotes *x* if and only if *x* is (a bit of) the element with atomic number 78 or *x* is (a bit of) the element with atomic number 79. Members of the two communities have different beliefs about what their word form 'gold' denotes, and they take these beliefs for granted even when they are evaluating utterances made by using gold in 1650. For instance, suppose that in 1650 John Locke and his twin on Twin Earth both uttered the words 'There are huge deposits of gold in those hills', with Locke indicating South African

hills and Twin Locke indicating the corresponding Twin South African hills, both of which contain platinum but no gold. We take Locke's word 'gold' to be true of an object *x* just in case *x* is gold, whereas our contemporaries on Twin Earth take Twin Locke's word 'gold' to be true of an object *x* just in case (as we would say it) *x* is either gold or platinum. We conclude that Locke's utterance is false, and our contemporaries on Twin Earth conclude that Twin Locke's utterance is true.<sup>16</sup>

### A DILEMMA FOR THE CAUSAL-HISTORICAL THEORY OF DENOTATION

Let's see if we can give a causal-historical explanation of our entrenched practical judgment that the denotation of the English word 'gold' did not change since 1650. Suppose that in 1650 members of both linguistic communities affirmed the following "ostensive definition":

(A) *x* is gold if and only if for most things *y* that I and other speakers in my linguistic community have on other occasions called gold, *x* is (a bit of) the same substance as *y*.

The question is whether any such ostensive definition in 1650 actually *determined* that gold is true of *x* if and only if *x* is (a bit of) the element with atomic number 79. The answer is "no."

To see why, note first that to explain our practical judgment that the denotation of the English word gold did not change since 1650 by appealing to (A), we must assume that:

(B) For all *x* and *y*, if *x* and *y* are gold, then *x* is (a bit of) the same substance as *y*.

is true in English and Twin English. Even if we stipulate that (B) is true in these languages, we have no good reason to believe that '*x* is (a bit of) the same substance as *y*' is true in *Twin* English of the ordered pair  $\langle x, y \rangle$  only if *x* has the same atomic number as *y*. The social (especially economic) practices in which the application of gold to gold or platinum is embedded in the Twin Earth community strongly suggest that if *x* is gold and *y* is platinum, for instance, then '*x* and *y* are gold' and '*x* is (a bit of) the same substance as *y*' are true in Twin English of  $\langle x, y \rangle$ , even though *x* does not have the same atomic number as *y*. One might try to rule this out by stipulating that:

(C) For all *x* and *y*, if *x* is (a bit of) the same substance as *y*, then *x* has the same atomic number as *y*.