

SOME REMARKS ON “EXTERNALISMS”

Tyler Burge

As far as I know, the term “externalism” is not used *prominently* by any of the philosophers on this panel to describe his own views.¹ The popularity of the term has been, I think, inversely proportional to its clarity. Like the term “naturalism,” “externalism” warmly welcomes a running together of different doctrines, perhaps in the interests of a felt solidarity among those in need of asserting allegiance to a cause. Much of the problem lies in unclarity about what is external to what. Some answers to such a question use terms like “meaning,” “proposition,” and “content” whose uses vary widely among philosophers. I want to make a few remarks about important differences among views that have been labeled “externalist.” I will lay aside epistemic externalisms and focus entirely on externalisms relevant to language and mind.

One family of views often termed “externalist” concerns language. It is trivial that many entities that are in fact objects of linguistic reference are external to – independent for their nature and existence of – language. It is *not* trivial that some factors *constitutively determining* which objects linguistic terms refer to are irreducibly external to – independent of – idiolects, dialects, and communal languages. Such factors include causal chains, contextual parameters, and molecular structures, none of which individuals need not be able to specify. This point was established, I think, by Strawson for demonstrative reference (Strawson 1959), by Saul and Keith Donnellan for proper names (Donnellan 1970; Kripke 1972), and by Saul and Hilary for natural kind terms (Kripke 1972; Putnam 1975b, 1975c). I certainly endorse the point. It marks one sort of view often termed “externalism.”

There are many importantly different and incompatible views about *meaning*, as distinguished from reference, that are called “externalist views.” A hyper-Millian view holds that the meaning, or the specifically semantical contribution, of names and perhaps other terms is *exhausted* by its referent. No one on this panel is committed to this position. There is the less committal but still Millian view that elements of linguistic meaning, or linguistic propositional content – particularly, elements of the meaning of proper

names and natural kind terms – *include* objects of reference in the physical environment. Hilary accepted such a view in the 1970s and Saul has shown some sympathy for it. I am non-committal, shading toward the doubtful.

I think that the notion of meaning has a root conceptual connection to the notion of potential understanding. I think that a direct connection between understanding and physical objects has never been made clear or plausible. Understanding is certainly always perspectival. The idea of understanding, or indeed perceiving, a physical object (even as a *component* of meaning) *neat* is, I think, incoherent. So if physical objects in the environment are to be considered components of meaning, some account of meaning that loosens its relation to understanding must be developed. Clearly, *some* sort of understanding, some sort of psychological competence, is associated with proper names and natural kind terms. A serious development of the fundamental notions of meaning and understanding that underlie even this moderate Millian view has never, to my knowledge, been undertaken. I believe that if semantical discussion is to get beyond the impasses that have faced it in the last couple of decades, more attention needs to be devoted to the conceptions and motivations that underlie various notions of meaning and understanding. So I believe that the moderate Millian view is at best unclear.

It does not help to add the negative point, originally made by Hilary, that for “externalists” meaning is not in the head. Insofar as meaning is shareable, even those who are labeled “internalists” can and should say that meaning is not in the head. I believe that meaning is an abstraction, hence not located anywhere. So I am doubtful that there is an acceptable notion of externality according to which it is true, interesting, and distinctive, to claim that *components* of meaning are external to language, the individual, or what not. But such a claim is commonly associated with the term “semantic externalism.” I think that the term is so associated with unclarity and dispute on these issues that it is better to drop it. “Semantic,” “meaning,” and “understanding” are sufficiently in need of explication that adding another term “externalism” to the mix just piles up intellectual debts.

In any case, I think that the power and interest of the arguments regarding linguistic reference that were given by Donnellan, Saul, and Hilary transcend technical issues about exactly how to regard meaning or linguistically expressed propositions. To return to something I said earlier, the constitutive *determiners* of reference are partly independent of individuals’ idiolects and of communal languages. Perhaps here we have a position that the three of us share that is very close to what is popularly understood as “externalism” about language.

Let us turn from language to mind. In my view, the positive results on linguistic reference are ultimately founded on a broader set of phenomena regarding mind. And the fact that the determiners of reference are external to idiolects, dialects, and communal languages is ultimately to be understood in terms of the independence of such determiners from individuals’ psychologies.

The term “externalism” is if anything less well anchored with respect to mind than it is with respect to language. Many associate the term with the view that representational mental contents or, even worse, mental states and events, are somehow in the wider environment beyond the individual. Although such fringe views haunt the landscape, no major figure I know of holds any such view. Some of Saul’s followers, though not Saul himself, maintain that whereas linguistic meaning includes or is exhausted by objects in the environment, representational mental content is purely descriptive. Thus referents of representational mental content are determined by description. Proponents of this two-dimensional view are sometimes counted “externalists,” although their position is “internalist” on the most elementary aspect, viz. reference, of the most fundamental matter, viz. thought. I leave perception aside here. The original arguments by Saul and Hilary (and by Donnellan) regarding linguistic reference can be trivially adapted to show that reference in thought cannot be fixed by descriptive capacities of individuals, or even whole communities. Since much reference in thought is not indexical, where reference is not fixed by descriptive capacities, representational thought content cannot be either. I am simply not generous enough to share an unqualified label “externalist” with such philosophers.

The view about mind that I provided a series of arguments for, beginning in the late 1970s, which I call “anti-individualism,” is that the constitutive determiners of the natures of many mental states include relations between the individual in those mental states and elements of the wider physical and social environments (Burge 2007). This is emphatically not primarily a view about representational mental content, but about mental states and events. It is about what representational mental content mental states or events constitutively have and can have.

The view is also not that the representational mental content is external to the individual. Again, I think of the content as abstract, and not anywhere. Moreover, *thinkings* of the content are always in the individual’s mind or head, if they are anywhere. (I think that claiming that there is an *exact* spatial location for higher-level mental events like thoughts may be pointless.) Thus externality to the individual concerns the *constitutive determiners* of the mental states’ representational content – and ultimately of the nature and identity of the mental state or event – *not* the content itself or the mental states or events themselves. The key point is that some constitutive determiners need not be part of or immediately accessible to the individual’s psychology. In this sense, elements beyond the individual – entities in the physical environment, and non-psychological, causal relations to them, for example – are among the constitutive determiners of the individual’s psychology. This view is not a trivial corollary of the work on linguistic reference.

Putnam’s work in “The Meaning of ‘Meaning’” (1975d) was at best ambivalent with respect to anti-individualism. But he came to endorse it fully. I think that it is the doctrine that underlies and helps explain the

views about linguistic reference that I characterized some paragraphs back. That is, I think anti-individualism is more fundamental than the views about linguistic reference that are often called “semantical externalism.” Language is an expression of mind.

“Externalism” about language should be firmly distinguished from “externalism” about mind. “Externalism” about constitutive determiners of the natures and identities of psychological (or mental) states or events should be sharply distinguished from “externalism” about the nature of meaning, representational content, and referents.

Even once such parameters are set, what externality precisely consists in is in many instances in need of clarification. I think that this label has encouraged sloppy thinking. Although proposals for linguistic reform are rarely heeded – and I do not expect this one to be – I believe that it would serve philosophy either to drop the term “externalism” altogether, or always to qualify it, so that the relata that are external to one another are clearly specified, and the particular relation of externality is clarified. Such usage would be less snappy. But short-hand has rarely served philosophical reflection.

Let me turn to a more broad-based observation about the currents in philosophy that I have been discussing. In retrospect, I am struck by the relatively compartmentalized focus on language that marked their beginnings. In the first half of the twentieth century, mainstream philosophy, especially in Britain and the United States, had focused on perception almost as much as on language. With the demise of sense-data views at mid-century, interest in perception abruptly faded. So the initial work on linguistic reference at mid-century was almost completely divorced from any discussion of perception.

Strawson is a partial exception to this remark. His work on demonstrative reference in the 1950s was linked to powerful points about the perception of duplicate scenes (Strawson 1959). He noted that a perception of a scene that is indiscernible from a perception that would be obtained from a similar scene that occurs elsewhere in the universe would nevertheless be a perception of, and only of, the scene by which it is causally occasioned. Here perceptual reference depends on causal relations to a scene and cannot be determined fully by discriminations available to the individual. In my view, Strawson failed to exploit this insight because of his descriptivist leanings with respect to names and his insistence on criteria for the application of general terms. In the early phases of the work on linguistic reference, no one besides Strawson thought hard about perception. Such was the tendency to reflect on language as an autonomous phenomenon.

Although this narrow focus on language reaped obvious rewards, I think that it also carried serious disadvantages. Sometimes perception was, and still is, regarded as dependent on language for making singular reference to entities in the physical environment, or as dependent on language for

categorizing entities under general types. Both ideas are completely out of step with what is known from science. More broadly, if a reasonable conception of perception had been central to these early discussions, the discoveries regarding linguistic reference and mental states would have come more easily. The discoveries about language would have been made in a broader and more natural setting. And deeper understanding of the origins of reference would have been attained.

For example, the Kripke–Donnellan points about reference of names have obvious analogs for perception. An object can be seen even though the perceiver could not perceptually distinguish the object – given the perceiver’s looking angle and background knowledge – from a look-alike that might have been substituted for the object. This point is analogous to the point that a language speaker can refer to an object with a name or to a kind by a natural kind term, even though the perceiver cannot distinguish the name’s or term’s bearer from other entities, except by use of the name or term. Perceptual reference is not carried out purely by perceptual attributives or perceptual categories in the perceiver’s (or perceptual system’s) repertoire, just as reference in language is not carried out purely by descriptive resources in the speaker’s repertoire. Perceptual reference, like reference with many names and kind terms, is partly and irreducibly determined by causal relations to the referent.

Similarly, perceptual reference is compatible with being perceptually wrong about most of the salient properties of the perceived object. The color, shape, sortal type, and position of an object can be misperceived, all at once, even as the object is perceived. Here we have a clear analog of a standard and deep point made about the use of names, natural kind terms, and so on. An individual can succeed in referring to the bearer of a name or to the referent of a natural kind term, even though most of the individual’s associated descriptions are false of the referent.

What was unfortunate about the narrow focus on language is not just a failure to exploit analogies between perceptual reference and linguistic reference. There was a failure to follow reference back to its roots. (I believe that this point bears on disquotationalist views of reference.) In many paradigmatic cases, the association of a name with its bearer, or of a natural kind term with the relevant kind, depends ultimately on perception. Elementary linguistic reference is, I think, to a large extent grounded in perception. There are forms of reference that are not obviously empirical – reference with numerals, for example. And there are empirical mechanisms beyond perception: chains of communication, introduction of kind terms through hypothesis or theory, and so on. But, obviously, the roots of reference for these more complex mechanisms, at least in empirical cases, lie in perception.

Even in my own work on mind, I did not center on perception until the early and mid-1980s, after I had studied David Marr’s psychology of vision in classes at MIT while I was a visiting professor there. Hilary does give

serious attention to perception in 1994 in his Dewey Lectures (Putnam 1994). In my view, however, these lectures do not make any serious use of the science of visual psychology – an area of psychology, almost the only area, that had matured well before then into a mathematically rigorous science. There is not a single serious reference to any of the methods or theories in empirical perceptual psychology in those lectures. In fact, much that Hilary says in those lectures seems to me to be incompatible with what had been known in the science for some decades.

A full conception of objective representation must develop accounts of mechanisms for linguistic reference in the light of a scientifically informed account of perception. I found it striking that in the roundtable on perception given at the Dublin conference – at least in the oral form that it took during the conference – there was not a single serious mention, much less extended discussion, of the considerable body of knowledge in the science of perception, particularly visual perception.²

Anti-individualism applies to perceptual states. Most of the anti-individualistic elements in thought and in language must be understood as grounded in antecedent perceptual representation. (I say more about these matters in Burge 2005, 2007, 2010.)

I turn now to a brief appreciation of Hilary's contributions, and some questions about them. I think of Hilary's work on reference, which really goes back to "It Ain't Necessarily So" (1975a), as brilliantly imaginative. The work is carried out with a nuance, depth, and freedom that make it entirely admirable. I think that it will prove to be one of his lasting contributions to philosophy. I would like to ask two questions about middle-level components of this work.

One is about stereotypes. Stereotypes were introduced, I think, in "Is Semantics Possible?" (1975b), as short dictionary-type explications of a term, anchored by a further component – the referent itself. I am primarily interested in the explication component of the stereotype. Such explications were accorded the function of conveying, in short-order, an understanding of the term that might get a novice started at applying it to its referents. Hilary rightly found remarkable the power of a short description to teach a language learner a new term, even though the description normally does not determine a unique referent, and normally is not even close to being synonymous with the term. He emphasized that the descriptive elements in the stereotype need not be *true* of all (or in the limit, perhaps any) of the entities to which the term applies, as long as those elements enable users to agree on a use with respect to standard or paradigmatic examples.

I regard this notion as an imaginative contribution to our thinking about meaning and reference. It is intriguing in that it does not fit into any of the standard slots in semantical thinking. I seriously doubt that the descriptive elements in stereotypes are, except perhaps in the loosest sense, any part of the literal meaning of the terms to which they are attached. They can

change with circumstance, context, and time in ways that terms' meanings do not seem to. The explicative, descriptive elements seem arguably a part of linguistic theory. They seem to be an important element in linguistic practice. But as I said, they do not seem to fit into any of the standard slots – literal meaning, implicature, or what not.

My question is how Hilary now regards this contribution – in particular, the explicative descriptive component of the stereotype – and whether he foresees ways of developing the notion further.

The other question is about twin-earth methodology. I owe much to Hilary's introduction of this methodology. I believe that it is an exceptionally imaginative and effective way of discussing examples. My attitude is not shared even by all who accept the conclusion of arguments that make use of the methodology. There are those who think that such dramatic fiction should play no role in philosophy, that the thought experiments are too strange to carry conviction, and so forth. I will not engage with wholesale doubts about the methodology, except to say that thought experiments using the methodology *did* carry conviction. I do believe that all the points made with the methodology can in principle be made without it. But discovery and imagination are often aided by just such thought experiments. They seem to be at least an epistemically valuable prompt for reflection.

What I want to focus on is a particular aspect of the thought experiments. In the literature subsequent to Hilary's publication of the initial twin-earth thought experiment, it was often noted that the thought experiment that centered on water describes a situation that is strictly impossible (Putnam 1975d). It seemed important to the thought experiment that it concern actual human usage. And it seemed important to the thought experiment that it concern a simple everyday kind like *water*. But one simply cannot have a duplicate twin of a human being or human body in a world in which there is no water.

From the moment the point was articulated, I believe that all philosophers with judgment, including Hilary of course, recognized that in some sense the impossibility did not matter to the point and force of the thought experiment. Even though the thought experiment rested on an impossible state of affairs, it made its point. How did it do so?

A possible view is that there are other thought experiments not involving water that postulate such duplications that are possible. But it seems that no thought experiment that takes a human being as one of the protagonists and that concerns the stuff *water* could possibly do the job. Deprive twin-earth of water and one cannot have a physical duplicate of a human being on twin-earth. Then why is it that the thought experiment is persuasive specifically regarding the referential relation between the term "water" and *water*?

It seems to me that the answer has to do with the point of the thought experiment. The main point is not what it first appears to be. The main point is not to show that it is metaphysically possible that physical

duplicates can have terms with different referents. The main point is not about the failure of local supervenience of linguistic reference on the physiology of the individual language user. This point can be made using other thought experiments. Hilary's water thought experiment does suggest how to construct such thought experiments.

But I think that the value of the twin-earth thought experiment that centers on *water* is not adequately accounted for by calling it a near miss. Its value is not best accounted for by saying that it fails on its own terms, but helps us think up better thought experiments that succeed. I believe that these reflections indicate that the point about the failure of local supervenience of linguistic reference on the physiology of individual language users is not properly thought of as the basic philosophical point (or objective) of Hilary's original thought experiment. The basic point is rather to bring to consciousness an awareness of what factors constitutively bear on reference, and on aspects of meaning that determine reference.³ The impossible fiction contains pointers to what factors are and what factors are not decisive in determining the nature of the referential relation, at least in the cases of a large class of kind terms. Even though strict duplicates are impossible, the postulation of the duplicates reminds us that what sorts of fluid are contained in an individual's body are not any more determinative of the referents of the individual's terms than are what sorts of connections obtain between the individual's synapses or what sorts of bodily movements the individual engages in, considered in themselves.

The science fiction makes a point that does not lie in a strict modal counter-example to the view that sameness of bodily constitution and behavior is compatible with difference in reference. The science fiction helps one reason more clearly about what factors are relevant and what factors are irrelevant to reference. I would like to understand better the relation between the fiction in the thought experiment – a fiction that makes its point even though it is quite literally an impossible fiction – and thought experiments that *depend* on citing a genuine possibility.

I am interested in what Hilary thinks about these matters. His use of science fiction in making philosophical points is unsurpassed. Perhaps there are epistemic insights embedded in his methodology that might help us better understand not only modal epistemology, but the epistemology of reflection on constitutive matters.

Notes

- 1 This piece is a lightly edited and lightly supplemented version of a talk given at a panel discussion whose designated title was "Externalism." The discussion occurred at the conference in honor of Hilary Putnam in Dublin, Ireland, March 2007. The other panelists were Saul Kripke and Hilary Putnam. I have retained first names for these colleagues in order to emphasize the informality of both my contribution and the occasion.

2 The chapters by Travis and McDowell in this volume.

3 The essay is advertised as being about meaning, but I believe that Hilary’s discussion of meaning in the essay is confused. I think that the main insights are into reference and derivatively into aspects of meaning that are different if reference is different. For some discussion, see my “Other Bodies” and Introduction in Burge 2007.

Bibliography

- Burge, Tyler (2005) “Disjunctivism and Perceptual Psychology,” *Philosophical Topics* 33: 1–78.
- (2007) *Foundations of Mind: Essays by Tyler Burge*, vol. 2, Oxford: Oxford University Press.
- (2007a) “Other Bodies,” in *Foundations of Mind: Essays by Tyler Burge*, vol. 2, Oxford: Oxford University Press, 82–99.
- (2009) “Perceptual Objectivity,” *Philosophical Review* 118(3): 285–324.
- (2010) *Origins of Objectivity*, Oxford: Oxford University Press.
- Donnellan, Keith (1970) “Proper Names and Identifying Descriptions,” *Synthese* 21: 335–358.
- Kripke, Saul (1972) *Naming and Necessity*, Cambridge, Mass., Harvard University Press.
- Putnam, Hilary (1975a) “It Ain’t Necessarily So,” in *Philosophical Papers*, vol. 1, Cambridge: Cambridge University Press, 237–49.
- (1975b) “Is Semantics Possible?,” in *Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press, 139–52.
- (1975c) “Explanation and Reference,” in *Philosophical Papers*, vol. 2, Cambridge, Cambridge University Press, 196–214.
- (1975d) “The Meaning of ‘Meaning’,” in *Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press, 215–71.
- (1994) “Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind,” *The Journal of Philosophy* 91: 445–517.
- Strawson, P.F. (1959) *Individuals*, New York: Doubleday and Company, 1963, ch. 1, sections 1–2.