Kripkean Externalism versus Conceptual Analysis

Abstract. Does Naming and Necessity show that we have no useful a priori knowledge about the reference of rigid designators in science and philosophy? Jackson and Chalmers answer No: Kripke shows only that the C-extension of a rigid designator is an empirical question, but any rigid designator has also its A-extension, and we know of it a priori. But what is A-extension? Supposedly, a dimension of extension which diverges from C-extension only at non-actual worlds, and only when these worlds are treated as being actual. This sounds contradictory. This paper argues that it does reflect a confusion—the confusion of seeing non-actual worlds as separated spatially, not modally, from the actual world. We have indeed no useful a priori knowledge concerning the reference of rigid designators. The externalism of Naming and Necessity does raise puzzles—but they are removed by Millikan’s naturalized account of concepts.

Does the externalism about meaning sketched by Kripke in Naming and Necessity (1972) really show that we have no useful a priori knowledge concerning the reference of many key terms in science and philosophy? Does it thereby show that there is little place or role for conceptual analysis? Frank Jackson (1998) and David Chalmers (1996) have recently argued that the answer to both questions is No. Kripke’s insights are compatible, Jackson and Chalmers maintain, with the idea that such key terms have two different kinds of extension—C-extension and A-extension—and that while a term’s C-extension is an empirical question, our knowledge of a term’s A-extension is indeed a priori.

This paper argues that the distinction between C-extension and A-extension is confused. For it requires us to think of the separateness of any one possible world from the others as spatial separateness, not modal. Indeed there is no good reason, this paper further argues, to suppose that we have any useful a priori knowledge concerning our key terms’ extensions. This uncompromising externalism may seem to create a mystery about how our rigid designators manage to refer as they do at all, and the paper closes by suggesting that Ruth Millikan may have dispelled the mystery.

I

The distinction between C- and A-extensions initially sounds clear enough. When asking about term T’s C-extension, Jackson writes, “we are considering, for each world w, what T applies to in w given
whatever world is in fact the actual world, and so we are, for all worlds except the actual world, considering the extension of $T$ in a counterfactual world” (p. 48). Now many of our key terms are rigid designators, as Jackson concedes Kripke to have shown. That is, in talk about any counterfactual world, they refer (if at all) to just the same stuff or kind or phenomenon as they do in talk about the actual world. So the C-extension of any rigid designator is, for all worlds $w$, the same as for the actual world. The C-extension of “water”, for example, is always $\text{H}_2\text{O}$ (p. 49; cf. Chalmers, p. 57, p. 59).

When we ask about a term’s A-extension, in contrast, “we are considering, for each world $w$, what the term applies to in $w$ given or under the supposition that $w$ is the actual world, our world” (p. 48). Thus consider a world in which (as Jackson puts it) “the watery stuff of our acquaintance” is, not $\text{H}_2\text{O}$, but XYZ. To ask about the A-extension of “water” in such a world is to ask what in that world “water” picks out “under the supposition that that world is the actual world” (p. 49). And the answer, Jackson maintains, is clear: the A-extension of “water” in that world is then XYZ (cf. Chalmers, pp. 57-58). In general, “where the kind common to the relevant watery examples…is kind $K$,” in any world $w$ regarded as actual, the A-extension of “water” in $w$ is kind $K$, “be that kind $\text{H}_2\text{O}$, XYZ, or whatever” (p. 49). We need not know whether $w$ is the actual world, nor what chemical kind populates the watery exemplars with which actually we are acquainted. So we can know what the A-extension of “water” is without empirical knowledge of the actual world. We can know it a priori (cf. Chalmers, p. 59).

But the distinction between C-extension and A-extension is less clear than it seems. Let me begin with a point about “water”’s C-extension. Why was it noteworthy for Kripke to observe that, in talk about any counterfactual world, “water” has the same referent (if any) as in talk about the actual world—i.e. to observe that “water” is a rigid designator? Because it corrected a tendency to suppose otherwise. In many counterfactual possible worlds which Kripke considers, we ourselves appear, speaking a language much like our actual one. In one we even wield a term which sounds just like “water”—but the samples of which we assert the term are composed of a chemical kind other than $\text{H}_2\text{O}$ (Lecture III passim, e.g. p. 124).
Does this possible world show that water might not have been H\textsubscript{2}O? There is a tendency to suppose so. But Kripke points out that the right way for us to describe such a possible world—situated as we are in the actual world, and speaking as we do our actual language—is to say that in such a world, the term of ours which sounds like “water” does not designate water. We must allow that there may be water in such a possible world. But we must add that it is found not in the samples by which, in that world, we teach the use of the term which sounds like “water”, and is instead found wherever and only where there is H\textsubscript{2}O in that world.

Does one variant of this possible world show that

(1) Our term “water” might have designated XYZ

is true? Only in a qualified sense. Our sound-parcel “water” might indeed have designated XYZ. But it does not follow that our term water—complete with its semantic attachments—might have designated XYZ. The only sense in which (1) is true is a sense too weak to license inference of the corresponding assertion in the material mode of speech, viz.

(2) The world’s water might have been XYZ.

So much by way of review. Now, what of “water”’s A-extension? Jackson’s definition reads thus: “the A-extension of \(T\) at \(w\) is the extension of \(T\) at \(w\) given \(w\) is the actual world” (p. 50). But the phrase “given \(w\) is the actual world” raises puzzles, largely owing to its employment of the indicative. After all, only one world \(\text{is}\) the actual world. True, then, that in the case of that one world, the application of Jackson’s definition is straightforward. But what about the other worlds, not one of which is the actual world? What are we to make of the instruction to seek out the extension of \(T\) in a world \(w\), which \(\text{is not}\) the actual world, given \(w\ \text{is}\) the actual world?

Jackson’s alternate wording says that we are to seek out the extension of \(T\) in \(w\), which is not the actual world, “under the supposition that \(w\) is the actual world”. But how are we to think of \(w\) as a world additional to the actual world—as itself \(\text{not}\) the actual world—while simultaneously supposing that \(w\ \text{is}\) the actual world?
My puzzlement may seem obtuse. It may seem obvious that when we apply Jackson’s definition to worlds other than the actual world, we are supposed to ask ourselves, “well, what if this other world had been the actual one—what if this other world were actual?”

But note that the use here of the conditional mood (or of the subjunctive) stamps the world of which we are speaking as counterfactual. So for worlds other than the actual one, Jackson’s definition of A-extension would, on this uncharitable interpretation, have us ask ourselves what the extension of \( T \) is at such a world, regarding that world as counterfactual. T’s A-extension will have collapsed into its C-extension.

So let us try a different interpretation: might the question about “water”’s A-extension be meant not as a question about worlds other than the actual world at all? Might Jackson and Chalmers really just be asking: “what if there is a single chemical kind found in our lakes and streams, but that kind is not after all H\(_2\)O, and is (amazingly) XYZ—what then is the extension of ‘water’?” But this reading seems even more uncharitable. For in that case Jackson and Chalmers are only reminding us that our judgements about C-extension are fallible, and may have to be revised. They have not pointed out a different form of extension.

But then just how might Jackson and Chalmers have supposed that rigid designators have a type of extension distinct from C-extension? I offer this suggestion. Think of the question about \( T \)’s C-extension, i.e. its extension in the various counterfactual worlds, as like a question about \( T \)’s extension in numerous spatially remote worlds. Standing here in the actual world, we train our telescope upon each world in turn, and determine where in each \( T \) is satisfied. But to ask about \( T \)’s A-extension is to ask about these other possible worlds, not as far away, but as rotated into the position occupied by the actual world—as replacing it, one after another, in the actuality slot. Small wonder that a different answer emerges.

Even this does not, I admit, seem a charitable conjecture. Possible worlds are removed from the actual world not spatially but modally, and our access to them is not via telescopes. And asking what to
say about an XYZ-containing world, *had it been* rotated into the actuality slot, is asking about the C-extension of “water”.

But there is evidence that my seemingly uncharitable conjecture is right, at least in Jackson’s case. In a footnote (p. 48, fn. 29), Jackson considers a puzzlement much like the one raised above: how can we entertain *seriatim* the hypotheses that world \( w_1 \) is actual, that world \( w_2 \) is actual, etc., when we know perfectly well that at most one world is actual? Jackson’s answer: imagine that you have been kidnapped and blindfolded, and at length hear your captors say “we are here at last”. Certainly you could entertain *seriatim* the hypotheses that the geographical location your captors term “here” is \( G_1 \), that the location is \( G_2 \), etc., even though you would also know that at most one of these locations can be truly termed “here” by them.

But the parallel Jackson draws here is deceptive—and revealing. Different sets of captors, simultaneously located at different geographical locations, can *actually and truly* assert “at last we are here”. (In a way, the kidnapped person’s task is to figure out just *which* such set of captors he is actually listening to.) But different claims as to which world is actual cannot *actually and truly* be made. For different claims as to which world is actual just are—in less mysterious language—divergent claims as to what *the* world is like. Suppose then it actually is claimed that *the* world contains in its streams and lakes—that is, in the samples of stuff used in teaching the term which has the sound “water”—\( \text{H}_2\text{O} \); and that it actually is claimed that *the* world contains in its streams and lakes \( \text{XYZ} \); and that it actually is claimed that *the* world contains there \( \text{H}_2\text{SO}_4 \); etc. Then at most one of these actual claims is true. All three such claims could indeed simultaneously be true, if made in different worlds. But that is because the separateness between such worlds is modal, not spatial. It is because only one of these worlds would be the actual world, and would contain an actual claim about the contents of streams and lakes.

Jackson (likewise Chalmers) wants us to regard as true more than one claim of the form “world \( w_x \) is the actual world”. It is indeed possible for truth to attach to a plurality of such claims, so long as it does not do so all at once—so long as it *actually* attaches to just one such claim, and *would* attach to the others in other possible worlds. That is, even once we allow that “world \( w_1 \) is actual” is true, we *can*
allow that “world w₂ is actual” is true as well, but only if we treat the latter just as a claim that would be made, not as an actual claim—or, equivalently, only if we word the latter as an actual claim, but in the conditional/subjunctive. But for it to be the case that more than one indicative claim of the form “world wₓ is actual” is true—is actually true—the worlds of which they are asserted must be separated spatially, not modally. That is after all how it can happen that more than one indicative claim of the form “location Gₓ is here” is true, and actually true: the locations of which they are asserted coexist at a remove from one another, rather than rival one another for existence. (Or, for Lewisians: the locations are co-actual at a remove from one another; they do not rival one another for actuality.)

II

Many will suppose that there must be something to the distinction between C-extension and A-extension—or to a distinction very much like it—since only such a distinction can articulate the sense that we do have some a priori knowledge concerning the reference of key rigid designators in science and philosophy. Concerning “water”, for example, it may seem that we at least know this a priori: it refers (if at all) to that kind of stuff, found in samples with which we are sensorily acquainted, which plays in our world “the watery role” (Jackson, p. 50). But this too is an illusion, I contend. A priori, we have no useful knowledge concerning the reference of any key terms in science or philosophy.

Note to begin with that there is no one “watery role” which we—we competent users of the world “water”—all believe water to play. True enough, just about all of us think of water as the clear, potable liquid which fills lakes and streams and falls to earth as rain. We can imagine exceptions: remote villagers who are surrounded by muddy lakes and polluted streams, and do wield the English word “water”, might think of water as only rarely clear and only sometimes potable. But such exceptions seem fanciful. More serious is the point that many of us associate with “water” richer descriptions—diverse richer descriptions, at that—than “clear, potable, and present in rain”. Many of us think of water as having a gaseous state, a liquid state, and a solid state, but there probably really are villagers who wield
“water” without having in their associated description “has a solid state”. Divers think of water as conducting sound better than air does; most of us do not. Some of us think of water as having an index of refraction of 1.3325. Many speakers of English indeed associate with “water” the description “has molecular structure H₂O”—many, but not all.

The differences in the descriptions which competent speakers associate with “water” come about because some of us have greater experience with water than others, greater empirical knowledge—not usually knowledge gained first-hand, to be sure, but empirical knowledge nevertheless. In the description which any one of us associates with “water”, at least most elements are properties which we have learned empirically to belong to water.

Still, one might want to ask, how can any of us know where to look, in order empirically to gain knowledge about water, unless we begin with a description which we associate a priori with “water”? One of the most interesting lessons of Naming and Necessity is that we can ask (and learn) about a given person or stuff or natural phenomenon while being almost wholly in the dark about its nature (Kripke 1972, pp. 84-85). Our beliefs about it may be stunningly incomplete, inaccurate even as far as they go, and they may be merely empirical. Remember the famous example (pp. 83-84): I may know virtually nothing about mathematics, may have been given by a friend a rudimentary explanation of what the Incompleteness Theorem is, and may use the description “discoverer of the Incompleteness Theorem” to help fix reference for my tokens of “Gödel”. I have just one individuating belief about Gödel. Does it amount to a priori knowledge? It arguably does not amount to knowledge at all; for all I know, Gödel plagiarized results from Schmidt. Or if I do know that Gödel discovered the Incompleteness Theorem, I know it empirically, on the evidence of my friend’s testimony.

The one case in which the way we fix reference affords a priori knowledge is the case of a designating expression introduced by explicit baptismal utterance. When Leverrier says “by ‘Neptune’ I shall mean the one planet responsible for [such-and-such] perturbations in the orbits of Uranus and Saturn”, we and he know a priori that Neptune—if there is such a planet—is responsible for those perturbations (p. 79n).
But Kripke nowhere says that all rigid designators are introduced into the language by baptismal ceremonies that are explicit in this way. For good reason: that claim would be very implausible. In order to frame a baptismal utterance I have typically to use other rigid designators. (Remember that even “yellow” may be a rigid designator of the property it picks out in the actual world (p. 128n.) And for my listeners to be prepared to employ my baptismal utterance, and put it together with what they know of the world, they must already have beliefs, the expressions of which will typically employ rigid designators. Baptismal utterance must presuppose that rigid designators are there in the language; it cannot solely by itself be responsible for their presence.

The proponent of conceptual analysis might rejoin that we know a priori at least this much about the reference of “water”: it refers (if at all) to just that natural kind which is found in a majority (at least) of the samples we point to in teaching the use of “water”. But there is no one standard collection of samples which “we” employ in teaching the use of “water”. Different subgroups of us use different samples; over time, some samples evaporate or drain away, and their place in language-instruction gets taken by other samples of watery-looking stuff; and there plainly is no a priori guarantee that a majority of the samples thus eventually employed belongs to any one natural kind. Should later speakers thus wander in their choice of samples, what follows is that they will have made wrong choices, not that “water” never referred.

III

Yet however intuitive the arguments in its favor, the picture of reference presented in Naming and Necessity may seem utterly mystifying. How is it that a person’s tokens of “Gödel” or “water” or “heat” can really refer to Gödel or water or heat, even though that person’s head is nearly empty of beliefs about the referent—perhaps entirely empty of true beliefs? Certainly the fact that the person’s tokens of the name or designator are copied from tokens uttered by other members of his speech community is part of the answer. But copied in what way, and from what other members? Though the text resists the
interpretation, Kripke sometimes is taken to have envisioned a *theory* of reference, on which the speaker’s tokens of a designating expression must derive from a causal chain that traces back to speakers whose baptismal utterances of “water” or “heat” were provoked by sensory contact with a sample or instance (Devitt 1996, pp. 163-71; cf. Devitt 1981). This “theory” does not dispel the mystery, but expands it. It seems magical that transmission of my tokens of “water” from baptizers long since dead can, even though untraceable by me or anyone else, confer upon my tokens a power to refer to just *that* chemical kind (cf. Evans 1973). Also, there are questions about how far the chain must trace back—to baptizers who spoke English, or Anglo-Saxon, or some older language?—and what sort of path it must take. Also, it is puzzling that their baptismal tokens of “water” managed to refer to water, given that reference *here* cannot have been underlain in the way the reference of *my* tokens is, according to the theory. Finally, it seems in the end undeniable that *something* suitable must be present or happening in my head, that somehow *reflects* the causal ancestry of my tokens of “water”, in order for my tokens really to refer to water. (After all, does Kripke not admit that if I token the name “Napoleon” with no intention to refer to any person to whom anyone else ever referred by it, my tokens do not refer to Napoleon?)

In this section I will offer a brief sketch of a response to the original mystery—the mystery of how tokens of a designating expression issuing from a head virtually devoid of accurate beliefs about a given kind or stuff or phenomenon can yet refer to that very kind or stuff or phenomenon. The response is not mine, and is not simple. Sketching it briefly may therefore be ill-advised, but also might be of interest. Here then is the response to this mystery found in Ruth Millikan’s *On Clear and Confused Ideas* (Millikan 2000). A head which issues tokens of “water” or “heat” or “tigers” does indeed have to possess something called a *concept* of water (or of heat or of tigers). But this *concept* is not, as one might suppose, an introspectible parcel of belief or of cognitive disposition, which grounds an ability to ask and learn about water or heat—to track them across different states of affairs in which they figure, to reidentify them. Rather the concept of, say, water just *is* the ability to reidentify it in different states of affairs (Ch. 1). This is important because an ability to reidentify can be cued to widely varying marks or features of the thing reidentified. Some such marks may attach to the thing only often, or may attach not
exclusively to that thing, and in this case the ability to reidentify will be fallible. Some may attach to it with perfect reliability but only in the actual world or even only under local circumstances, and in this case the ability to reidentify may fail us in considering unfamiliar circumstances or other possible worlds (Ch. 4). But the most important consequence flows from the fact that the kinds and stuffs and phenomena of nature typically are characterized by a range of distinctive features or marks. It follows that different speakers may have an ability to reidentify the same kind or stuff or phenomenon but may do so by different means (Ch. 1, Ch. 5). You may be able to detect lemons by their fragrance or their texture, and not just by their taste or their visual appearance under standard lighting conditions. Thus different speakers may harbor different beliefs about a given kind (or stuff, etc.), and different dispositions to coin new beliefs about members, while yet all alike have an ability to reidentify that kind. The speakers then all have a concept of that kind (etc.), in Millikan’s parlance, but have different conceptions of it (Ch. 3, Ch. 4).

Millikan’s view thus preserves Kripke’s insight that the descriptions which a speaker associates with a given designating expression may be only contingently true of the referent. It also incorporates the point that different speakers may, by the same designator, refer to the same stuff, though they associate with that designator different descriptions—different versions of “the watery role”, for example. But what of the point that a speaker may refer to a given kind (or stuff, etc.) even though virtually devoid of beliefs about it—perhaps wholly devoid of true beliefs? Can a head so empty harbor an ability to reidentify the kind or stuff or individual in question?

Millikan might begin a response to this challenge, it may seem, by pointing to the case of perceptual tracking. To track an object perceptually is to reidentify it across different locations, under different perspectives and lighting conditions, perhaps in different postures or conditions. Such reidentification can be accomplished by a head which harbors no discursive beliefs at all about the object tracked—for even animals accomplish such reidentification. (I will say nothing about Millikan’s detailed discussion, in Ch. 13, of Gareth Evans.) Yet this answer seems unlikely to get very far. We need an answer on how a head nearly empty of information about Aristotle can reidentify Aristotle across
different states of affairs, even though Aristotle is no longer around to be perceptually tracked, or how a similarly empty head can reidentify electrons even though they are too small for perceptual tracking, or how one can reidentify heat even though what our senses directly track is rather the sensation produced in us by heat.

Here Millikan makes a deeply original move (Ch. 6). I will present it in two steps. First, she argues (on independent grounds) that “seeing Dan Rather on television”, or “seeing a microbe through a microscope”, are truly cases of seeing, in all philosophically relevant respects, even though they yield no non-inferential beliefs about one’s spatial relationship to the thing seen; “hearing one’s son’s voice on the phone” similarly is truly a case of hearing. Secondly, she argues that to learn about Dan Rather or microbes through the reports of one’s fellow speakers is, in all philosophically relevant respects, to perceptually track Dan Rather or microbes in yet another medium—not television or the microscope, but linguistic reports. In the historically normal case, she argues, hearing is believing, much as seeing is normally believing: an indicative report about the world received from a fellow speaker goes straight into belief, without mediation of inference by the listener concerning the speaker’s intentions and veracity. Moreover, in the historically normal case, the fellow speaker’s report is true. (As will be apparent, “historically normal” does not mean “statistically average”.)

Does this mean that a speaker whose head is nearly devoid of beliefs about electrons or Aristotle need only believe what fellow speakers tell him about those things, for it to be true that he himself can reidentify them, and hence that his tokens of “electron” and “Aristotle” can refer to them? If so, perhaps we have captured what seemed elusive but true in Kripke’s observation that the reference of the uninformed person’s tokens of “Gödel” connects crucially with the fact that they are copied from tokens uttered by his fellow speakers. But the ability to reidentify electrons or Aristotle involves just a bit more, for Millikan, than we have so far said—and this “more” validates our earlier sense that the ancestry of my tokens of “Gödel” must be complemented by something in my head.

Just what does it take for it to be true of someone that he can reidentify, say, squirrels? Suppose the person is largely empty of beliefs about squirrels, in much the way Kripke’s man is empty of beliefs
about Gödel. Our person does not know what squirrel tracks look like, how squirrels’ vocalizations sound, etc. He knows a squirrel to see one, and can do short-term perceptual tracking. But when we ask him to find the unusually fat squirrel we all saw ten minutes ago, he is as likely to look up in the sky, down in the pond, or in his checkbook, as at the trees. Such a person does not in any full sense have an ability to track squirrels across states of affairs. For that, he would need to know where to look to find the squirrel again.

So too, Millikan says, in the case of what one might call “conceptual tracking”—tracking squirrels across states of affairs perceived in the medium of linguistic reports. To do that, one needs to know where to look—i.e., which reports to elicit—to have a true ability to reidentify. One needs to know what sorts of questions to ask to learn how matters stand with, say, squirrels, or gold, or people, or heat. One needs to know that people have sizes and shapes, and that heat and gold (as opposed to samples of gold) do not; that heat comes in different degrees, but people and gold do not; that gold occurs in samples of different size and weight, while heat does not, and a person does not. To have the concept of gold, then, one does not need to have any interesting true beliefs about what gold is like—but one does need to know in rough outline what sorts of things there are to be learned about it (Ch. 5). In Millikan’s parlance, to have a concept of a given “substance”—that is, of a given kind or stuff or individual—one needs to wield a “template” outlining the sorts of properties which respect to which that substance is stably and determinately characterized.

But this may raise an alarm: does grasp of such a template amount to a priori knowledge about the kind or stuff in question, indeed a priori knowledge rather like the knowledge which Jackson and Chalmers find in our grasp of A-extensions? The answer is No. Empirical findings can in principle invalidate any aspect of a given substance template (p. 30). I may begin by supposing lemons to be stably determinate with respect to color, but may then learn from experience with Indian lemons that color does not belong to the relevant template. I nevertheless learn of lemons that they vary with respect to color: my concept continues to be a concept of the same kind, though my reidentification of that kind comes to
be cued to different sorts of properties from before; for ability-to-reidentify typically is cued to diverse and shifting features of the kind (or stuff, etc.) that is reidentified.

I shall end my sketch here. The point which I intend the sketch to make plausible, at least, is this. The ability to reidentify a kind or stuff or phenomenon, hence to refer to it by tokens of a designator, need not go with any positive or substantive a priori knowledge about the item reidentified—nor even with readily articulated empirical knowledge about it. That ability is rather an implementation of programs instilled in us by natural selection for attuning our thoughts and words to the world. How such programs work may no more be open to a priori introspection than are the workings of any other capacities installed in us by natural selection. Naming and Necessity can be de-mystified in an externalist way. It is unnecessary to graft onto it an internalist, and ultimately verificationist (cf. Millikan, p. 72), dimension of meaning. But that is precisely what bi-dimensionalism does, I believe. I urge that we consider abandoning it.
1 This sort of claim is not original to Jackson or Chalmers; other versions are found in Martin Davies and I. L. Humberstone (1980), in Pavel Tichy (1983), and elsewhere besides. What is distinctive of Jackson and Chalmers is the contention that, of the “two dimensions” of extension, one is a subject of a priori knowledge; see Block and Stalnaker (1999).

2 Chalmers’ wording has it that we are to consider the extension of $T$ in the various non-actual worlds “considered as actual” (1996, p. 60). This seems no better.

3 He holds that reference-fixing for terms that designate species is like the reference-fixing that occurs in a “baptism”, but that it would be “artificial” to (p. 135) to suppose there literally is a “baptism”. Cf. pp. 138-39 and p. 162.

4 Kripke, pp. 138-39, and p. 139n; I take these passages as correcting what is said at the top of p. 136.
References


Devitt, Michael 1996: *Coming to Our Senses*. Cambridge: Cambridge Univ. Press.


