Folk semantic intuitions, arguments from reference and eliminative materialism

Mark Pinder
University of Hertfordshire
mark.jonathan.pinder@gmail.com


Abstract. In a series of papers, Machery, Mallon, Nichols and Stich critique so-called arguments from reference, arguments that assume a theory of reference in order to establish substantive conclusions. The critique is that, due to cross-cultural variation in semantic intuitions giving rise to methodological problems in the theory of reference, all arguments from reference have an unjustified assumption. I examine an important example of an argument from reference, an argument of Churchland’s in support of eliminative materialism. I suggest that extant responses to the critique are unsatisfactory, and provide an alternative response: one might justify the assumption of a theory of reference in an argument from reference by appealing to an appropriate explication of the relevant commonsense concepts.


1. Introduction

According to Edouard Machery, Ron Mallon, Shaun Nichols and Stephen Stich (henceforth MMNS), arguments from reference are “arguments that derive philosophically significant conclusions from the assumption of one or another theory of reference” (Mallon et al. 2009: 332).

Here, a theory of reference is a theory that seeks to provide a systematic basis upon which worldly items are assigned to some collection of words and phrases, or to the concepts they express. We will be concerned with two principal types. First, descriptivist theories of reference hold that competent speakers associate reference-fixing descriptions with the relevant terms or concepts, and that the referent of a given term or concept is whatever satisfies, or comes sufficiently close to satisfying, the relevant reference-fixing descriptions. Second, causal-historical theories of reference hold that a term or concept t refers to an entity x just in
case there has been an appropriate causal chain of users acquiring $t$ from other users, such that the chain began with an initial ‘baptism’ of $x$.

According to MMNS, arguments from reference can be found “in nearly every corner of philosophy, including the philosophy of mind, the philosophy of science, the philosophy of race, and meta-ethics” (Mallon et al. 2009: 332). Herein, I focus on one of their principal examples: an argument offered by Paul Churchland (1981) in defence of his eliminative materialism.

Eliminativists propose that mental state terms like ‘belief’ and ‘desire’ are defined by their role in […] the folk theory of mind. They assume that if these terms have referents, they must be entities that satisfy (or come close to satisfying) the relevant definitions. That is, eliminativists assume some version of a descriptivist theory of reference for mental state terms like ‘belief’ and ‘desire’. […] Eliminativists claim that the emerging scientific facts suggest that nothing satisfies the descriptions folk psychology associates with ‘belief’ and ‘desire’. Thus, ‘belief’ and ‘desire’ do not refer [and] beliefs and desires do not exist (Mallon et al. 2009: 334).

According to MMNS, then, Churchland assumes a descriptivist theory of reference to derive the philosophically significant conclusion that beliefs and desires do not exist.

MMNS present a general critique of arguments from reference (Machery et al. 2004, 2013; Mallon et al. 2009). The critique proceeds in three stages. First, MMNS present experimental data that, they believe, provide evidence of cross-cultural variation in semantic intuitions—and, in particular, cross-cultural variation in intuitions about reference.

In two separate studies […], we found that Americans were more likely than Chinese to give causal-historical responses. […] As we had predicted, Chinese participants tended to have descriptivist intuitions, while Americans tended to have [causal-historical] intuitions (Mallon et al. 2009: 341)

Second, MMNS argue that such cross-cultural variation undermines the standard methodology used in theorising about reference. They characterise the methodology as follows:

Although philosophers have rarely been explicit about what the correct method is for determining the right theory of reference, they typically appeal to the intuitions of competent speakers about the reference of proper names (or other kinds of words) in actual and possible cases. […] In practice, philosophers usually appeal only to their own intuitions about reference and those of a few colleagues, perhaps because they take these intuitions to be representative of competent speakers’ intuitions or perhaps because they take them to be more reliable. (Machery et al. 2013: 620)
And they raise a problem for this methodology along the following lines:

[T]he evidence suggests that it is wrong for philosophers to assume a priori the universality of their own semantic intuitions. [...] We find it wildly implausible that the semantic intuitions of the narrow cross-section of humanity who are Western academic philosophers are a more reliable indicator of the correct theory of reference (if there is such a thing [...] ) than the differing semantic intuitions of other cultural or linguistic groups (Machery et al. 2004: B8–B9)

Third, MMNS claim that, without an acceptable methodology for theorising about reference, the assumption of any particular theory of reference is unjustified. As such, every argument from reference has an unjustified assumption. So, according to MMNS, all arguments from reference fail.

The general critique has given rise to a great deal of discussion, which has focused around the first and second stages. In light of the first stage, theorists have examined whether the available data do in fact support the conclusion that there is cross-cultural variation in semantic intuitions; and in light of the second stage, theorists have examined whether cross-cultural variation would in fact undermine the methodology involved in theorising about reference. As things currently stand, further experiments have suggested that the evidence in favour of cross-cultural variation is reasonably robust, but substantial disagreement remains about the consequences for methodology.

Such responses to the critique, however, may do little to satisfy proponents of arguments from reference. For example, Churchland seeks to conclude that there are no beliefs, desires, etc. Importantly, this is intended to be a substantial conclusion about human psychology. Thus, he states that

our common-sense conception of psychological phenomena constitutes [...] a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. (1981: 67)

Such a conclusion, if correct, ought to be independent of whether (a) there is the kind of cross-cultural variation in semantic intuitions emphasised by MMNS, or (b) such cross-cultural variation would undermine methodology for theorising about reference. It is not plausible to

---

1 For critical discussion of the first stage, see e.g.: Deutsch 2009: 453–457; Lam 2010; Martí 2009; and Sytsma and Livengood 2011. For responses, see (respectively): Machery et al. 2015; Machery et al. 2010; Machery et al. 2009; and Sytsma et al. 2015. For discussion of the second stage, see e.g.: Andow 2014; Deutsch 2009; Devitt 2011, 2012; Ichikawa et al. 2012; Machery 2012; and Machery et al. 2013. For a recent collection of papers, see Haukioja 2015 and, for a more general introduction to the literature, see e.g. Genone 2012 and Hansen 2015.
think that such factors as (a) and (b) tell us so directly about mental architecture. What Churchland needs, then, is a response to MMNS’s critique that does not take a stand on such issues. In particular, he needs to provide an alternative strategy for justifying the assumption of a descriptivist theory of reference in his argument from reference: a strategy that does not rely on anything like MMNS’s ‘standard’ methodology.

The point generalises. All of MMNS’s examples of arguments from reference seek to establish substantial conclusions—that scientific knowledge does (or does not) increase, that race is (or is not) real, and that ethical knowledge is akin to scientific knowledge—which, on the face of it, should be independent of (a) and (b). In order to assess the general force of MMNS’s critique, we might thus aim to establish whether or not there is a general strategy for justifying the assumption of a theory of reference in arguments from reference—a strategy that does not rely on anything like MMNS’s ‘standard methodology’. The principal aim of this paper is to motivate and develop a strategy—which we may call the explication strategy—along these lines.

As already indicated, I focus throughout on Churchland’s argument from reference. The reason for this is twofold. First, as the devil will inevitably be in the detail, it is instructive to see how the details of the explication strategy can be worked out in at least one particular case. Churchland’s argument from reference is an ideal case if only because it is reasonably well-known. Second, as I argue below, there is at least one way of interpreting Churchland’s argument from reference on which the explication strategy provides a natural response to MMNS’s critique. Spelling this out will serve as a motivation for the explication strategy more generally.

2. Churchland’s argument from reference

Churchland’s argument from reference is offered in favour of his eliminative materialism, the view that the common-sense mental-state concepts (of belief, desire, etc.) are theoretical concepts of an empirical theory, folk psychology, that is radically false and whose ontology will ultimately be replaced. Its structure is as follows.

(P1) The common-sense mental-state concepts are theoretical concepts of folk psychology.

Contemporary science shows that it is a serious possibility that folk psychology is radically false.

It is a serious possibility that there are no beliefs, desires, etc.

Here, folk psychology is to be understood as a system of laws, amongst which are those that connect up external conditions with mental states (such as (1)), those that govern mental states (such as (2)), and those that connect up mental states with actions (such as (3)).

1. For all $x, y$: if $x$ is facing $y$ and $y$ is unobstructed, then (ceteris paribus) $x$ is aware of $y$.

2. (i) For all $x, p$: if $x$ fears $p$, then (ceteris paribus) $x$ desires not-$p$.
   (ii) For all $x, p, q$: if $x$ believes $p$ and $x$ believes [$p$ implies $q$], then (ceteris paribus) $x$ believes $q$.

3. For all $x, p, \varphi$: if both $x$ desires $p$, and $x$ believes that $\varphi$ing will bring about that $p$, then (ceteris paribus) $x$ will $\varphi$.

The theoretical concepts of folk psychology are just the concepts expressed by the non-logical vocabulary in such laws.

As mentioned above, Churchland’s argument assumes that the theoretical concepts of folk psychology have their referents fixed descriptively, with the laws of folk psychology acting as the reference-fixing descriptions. It is only with this assumption that the radical falsity of folk psychology suffices to establish that there are no beliefs, desires, etc. (For example: our uses of the theoretical terms of folk psychology might stand at the end of appropriate causal-historical chains, at the other end of which are successful baptisms, even if the laws of folk psychology are radically false of whatever was baptised.)

It is worth noting that the argument does not assume what we might call a strict descriptivist theory of reference. That is, the argument does not assume that, for the theoretical concepts of folk psychology to have referents at all, the reference-fixing descriptions must be perfectly satisfied. The strength of (P2)—which suggests that folk psychology may well be radically false—allows for some leeway: the reference-fixing descriptions need only be satisfied to a sufficiently high degree. Thus Churchland’s argument assumes that the concepts of folk psychology will refer so long as there are some internal states that come sufficiently close to satisfying the laws of folk psychology—whether or not they satisfy those laws
perfectly. The alleged radical falsity of folk psychology is then intended to warrant the inference that the theoretical concepts of folk psychology are empty.

Churchland summarises his defence of (P1) as follows.

Seeing our common-sense conceptual framework for mental phenomena as a theory brings a simple and unifying organization to most of the major topics in the philosophy of mind, including the explanation and prediction of behavior, the semantics of mental predicates, […] the other-minds problem, the intentionality of mental states, the nature of introspection, and the mind-body problem. Any view that can pull this lot together deserves careful consideration. (1981: 68).

I will discuss some of these aspects of the ‘simple and unifying organisation’ presently. Churchland defends (P2) by arguing

that folk psychology suffers explanatory failures on an epic scale, that is has been stagnant for at least twenty-five centuries, and that its categories appear (so far) to be incommensurable with or orthogonal to the categories of the background physical science whose long-term claim to explain human behavior seems undeniable. (p. 76)

As (P2) will not concern us in what follows, I put Churchland’s defence of it aside. See his 1981 (pp. 72–76) for more details.

Focus now on Churchland’s defence of (P1). Notice immediately that Churchland offers us neither a linguistic argument, nor a form of conceptual analysis; he does not provide linguistic evidence that “belief”, “desire”, etc., act like theoretical terms, nor does he seek to analyse in any obvious sense the everyday concepts of belief, desire, etc. Rather, he points to certain theoretical benefits obtained by thinking of the familiar mental-state terms and concepts as the theoretical terms and concepts of folk psychology.

As Churchland is typically understood, his specific aim here is to argue that the folk in fact possess the theoretical concepts of folk psychology—deploying them when commonsensically explaining or predicting actions, and expressing them with ordinary, run of the mill uses of “belief”, “desire”, etc. However, I want to suggest an alternative interpretation of Churchland’s defence of (P1), as a kind of Carnapian explication.³ To be clear, the suggestion is not that, somewhat surprisingly, Churchland has always intended his argument to be interpreted as a kind of Carnapian explication. Rather, the suggestion is that the specifics of

³ See Carnap 1950. I note in advance that nothing that follows depends on the details of Carnap’s ‘metaontological’ views.
Churchland’s argument, if not the surrounding text, lend themselves to the alternative interpretation. And, as we will see, interpreting Churchland’s defence of (P1) as a kind of Carnapian explication will help us respond to MMNS’s critique of arguments from reference.

Let me say a little more about explications. I take an *explication* to be a stipulated refinement, or precisification, of certain pretheoretic (or otherwise imprecise) concepts, typically to facilitate subsequent theorising. Following Carnap, I will call the refined concept (or the corresponding term) the *explicatum*, and the pretheoretic or imprecise concept (or the corresponding term) the *explicandum*.

Explications are *stipulative*: an explicatum is not intended to *describe* the explicandum, nor to make explicit its satisfaction conditions or encode its ordinary usage, etc.; rather, the explicatum is offered as a theoretical counterpart to the explicandum. When theorising, so goes the thought, one *ought* to use the explicatum *rather than* the explicandum. Thus, while explications are not true or false, there may nonetheless be good explications and bad explications. Whether an explication is good will depend on a variety of factors. For Carnap, the following three factors are the most important.

(I) The explicatum should be precise.
(II) The explicatum should be similar in relevant respects to the explicandum.
(III) The explicatum should be a fruitful concept.

A few brief comments. I interpret (I) as telling us that the more precisely defined an explicatum, the better it is.4 Following Carnap, I take the *similarity* demanded by (II) to be loose, and I take

---

4 An anonymous referee for this volume points out that this criterion is ‘risky’, as we can make an explicatum more precise by stipulating a theory of reference for it (a claim I rely on below). The issue arises with the stipulation of a *descriptivist* theory of reference: the more precise the reference-fixing descriptions, the more likely it is that our descriptions will go unsatisfied and our explicatum will be empty. In part because of such concerns, I suspect that criterion (I) should ultimately be amended: we would typically like an explicatum to exhibit an ‘appropriate level’ of precision relative to our background theory and knowledge. For example, the explicatum given below for *planet* is certainly not maximally precise, in part because there is no precise account of when a celestial object should count as having cleared its orbit of debris. But, until we have developed theories that rely on particular precisifications, or until we have observed the orbits of a greater range of candidate planets, it would be undesirable to arbitrarily precisify the explicatum: relative to our current cosmological theory and knowledge, the explicatum of *planet* is appropriately precise—and, in that particular case, the appropriate level of precision is not ‘risky’ in the relevant sense. Regardless, as the principal aim of this paper does not rely on the particular choice of criteria for a good explication, for present purposes I adopt Carnap’s criterion (I) as it stands.
the relevance to be dependent on the specific explicandum; I will deem (II) to be satisfied if the theoretically relevant, salient features of the explicandum are shared by the explicatum. And, following Philip Kitcher’s treatment of explication (2008: 115), I take the fruitfulness required by (III) to be a matter of how many theoretical questions and problems the explicatum helps us to resolve, perhaps weighted by their importance.

For example, consider the concept of *planet*. Until recently, there was no agreed upon definition—merely nine canonical instances. Then, in 2006, the International Astronomical Union explicated the concept, refining it so as to provide a better taxonomy of celestial objects. A *planet* was henceforth to be an object such that: it orbited a star but did not orbit another planet; it was large enough for gravity to have formed it into a sphere but not large enough for its gravity to trigger fusion; and it had cleared its orbit of debris. Here, the explicandum is the pre-2006 concept, with its nine canonical instances; and the explicatum is the refined concept with its tripartite satisfaction condition.

The explication of the concept *planet* plausibly satisfies (I)–(III). First, given that the explicatum is more precise than the explicandum, we might plausibly claim that (I) is satisfied. Similarly, (II) is plausibly satisfied: the most salient features of the explicandum are its nine canonical instances, eight of which—all except Pluto—fall under the explicatum. Finally, insofar as the explicatum provides a more unified cosmological kind than the explicandum, the explicatum is more likely to facilitate successful problem solving. So we might plausibly hope that, in due course, we will find that (III) is also satisfied. Plausibly, then, the IAU’s explication of the concept *planet* is a good explication.

Turn back to Churchland’s defence of (P1). We can interpret that defence as seeking to justify the following:

(4) The common-sense mental-state concepts (such as *belief*, *desire*, etc.) are to be explicated, and the explicata are the theoretical concepts of folk psychology.

To see how this works, let me go through Churchland’s discussion of three aspects of the ‘simple and unifying organisation’ that are brought to the philosophy of mind: the semantics of mental predicates; the explanation and prediction of behaviour; and the other-minds problem.

---

5 I roughly follow Ludlow’s (2014: 41ff) discussion here, although Ludlow is concerned with meaning modulation rather than explication.
First, consider the semantics of mental predicates. Churchland states that

the semantics of the terms in our familiar mentalistic vocabulary is to be understood in the same manner as the semantics of theoretical terms generally: the meaning of any theoretical term is fixed or constituted by the network of laws in which it figures. (1981: 69)

One key thought here is that we can straightforwardly give a semantics for theoretical terms, perhaps in terms of the network of laws in which those terms figure. This suggests that those theoretical terms, and the concepts they express, are reasonably precise, in the sense that we can provide conditions for correct application. For Churchland, then, the theoretical terms and concepts of folk psychology, qua explicata, go some way towards satisfying (I).

Second, consider the explanation and prediction of behaviour.

The fact is that the average person is able to explain, and even predict, the behavior of other persons with a facility and success that is remarkable. Such explanations and predictions standardly make reference to the desires, beliefs, fears, intentions, perceptions, and so forth, to which the agents are presumed subject. But explanations presuppose laws—rough and ready ones, at least—that connect the explanatory conditions with the behavior explained. The same is true for the making of predictions [...]. Reassuringly, a rich network of common-sense laws can indeed be reconstructed from the quotidian commerce of explanation and anticipation; its principles are familiar homilies; and their sundry functions are transparent. Each of us understands others, as well as we do, because we share a tacit command of an integrated body of lore concerning the law-like relations holding among external circumstances, internal states, and overt behavior. (1981: 68–69)

At the beginning of this passage, Churchland points to a distinctive feature of the common-sense mental-state concepts, viz. that they are deployed in our everyday explanations and predictions of behaviour. Such explanations and predictions, Churchland claims, presuppose (perhaps rough and ready) laws that connect up the ‘explanatory conditions’ with the ‘behaviour explained’. In the present case, the presupposed laws will connect up: (a) the conditions under which our ordinary explanations and predictions of behaviour are given; with (b) the relevant actions of other people. As is made clear at the end of the passage, the laws that underpin our everyday explanations and predictions of behaviour connect (a) and (b) via internal states.

There are two important ideas here: first, it is a distinctive feature of the common-sense mental-state concepts that they are deployed in everyday explanations and predictions of behaviour; and, second, such explanations and predictions are underpinned by a ‘tacit
command’ of laws that connect up external conditions with internal states and that connect up internal states with actions.

Folk psychology reflects these two ideas. First, the theoretical concepts of folk psychology feature in a set of laws—such as those in (1)–(3)—that underpin explanations and predictions of behaviour comparable to our everyday explanations and predictions of behaviour. As such, the distinctive feature of the common-sense mental-state concepts—that they are deployed in everyday explanations and predictions of behaviour—is reflected by the theoretical concepts of folk psychology. Second, the laws that underpin such explanations and predictions do not tie (a) to (b) directly, but do so via internal states. Thus, in this respect, the theoretical concepts of folk psychology, qua explicata, go some way towards satisfying (II).

Third, consider the other minds problem.

The problematic conviction that another individual is the subject of certain mental states is not inferred deductively from his behavior, nor is it inferred by inductive analogy from the perilously isolated instance of one’s own case. Rather, that conviction is a singular explanatory hypothesis of a perfectly straightforward kind. Its function, in conjunction with the background laws of folk psychology, is to provide explanations/predictions/understanding of the individual’s continuing behavior […]. Knowledge of other minds thus has no essential dependence on knowledge of one’s own mind. Applying the principles of folk psychologising to our behavior, a Martian could justly ascribe to us the familiar run of mental states, even though his own psychology were very different from ours. (1981: 69)

As I understand Churchland, he here takes the key challenge posed by the problem of other minds to be to explain how knowledge of other minds (i.e. knowledge that another agents believes, desires, etc.) is in principle possible—without appealing to ‘deduction from behaviour’ or ‘induction from the perilously isolated instance of one’s own case’. Now, from the perspective of folk psychology, it is natural to construe ascriptions of mental states as explanatory hypotheses. And, importantly, such hypotheses can be confirmed if the explanations and predictions they provide are sufficiently successful.6 Thus, from the perspective of folk psychology, one plausibly obtains the following important consequence.

---

6 For Churchland, of course, those explanations and predictions are in fact not sufficiently successful. The point is that, in principle, were those explanations and predictions sufficiently successful then one could come to know that another agent has beliefs and desires.
In principle, our everyday explanatory and predictive practices could lead to knowledge of others’ minds (i.e. knowledge that another agent believes, desires, etc.).

Thus, in this sense, the theoretical concepts of folk psychology are fruitful: they facilitate the resolution of theoretical problems in the philosophy of mind. That is, they plausibly satisfy (III).

Other aspects of Churchland’s defence of (P1) can be construed similarly—in particular, as supporting the conclusion that the theoretical concepts of folk psychology, qua explicata, satisfy (III)—although for brevity I do not discuss those aspects here.

Churchland’s defence of (P1) is naturally interpreted as a defence of an explication of the common-sense concepts of belief, desire, etc., viz. that they should be explicated using the theoretical concepts of folk psychology. If the argument (so construed) is right, then for theoretical purposes we should henceforth treat such terms as “belief”, “desire”, etc., as theoretical terms of folk psychology.

Given this explication, we can understand (P1) so that it follows trivially. If Churchland is within a theoretical context, then he should treat the common-sense mental-state concepts as the theoretical concepts of folk psychology: the latter act as straightforward replacements for the former. The premise (P1) can be interpreted as making explicit this commitment.

To be clear, MMNS’s critique does not extend to (P1), so construed. As I have characterised the premise, it is justified by an explication. Here, however, explications are stipulative: a good explication is not in any sense ‘true’, need not adhere to everyone’s intuitions, and, in particular, does not depend on an acceptable methodology for theorising about reference. One would not undermine this defence of (P1) by presenting empirical evidence that there is cross-cultural variation in intuitions about theoretical terms or about the concepts of belief, desire, or anything similar.

3. A response to MMNS’s critique

We can understand (P1) as making explicit an explication—where explications are not affected by the kind of methodological concerns raised by MMNS. As such, if we interpret Churchland’s argument as I have suggested, the following strategy for responding to MMNS’s critique emerges: one might subsume the assumption of a descriptivist theory of reference under (P1). That is, one might begin by replacing (P1) with the following:
The common-sense mental-state concepts are (to be explicated by the) theoretical concepts of folk psychology whose extensions are fixed descriptively.

Then, one might extend the original defence of (P1) to support the new (P1'). In particular, one might argue for an explication of the common-sense concepts of belief, desire, etc., such that the explicata are the theoretical concepts of folk psychology construed descriptively. As explications are stipulative, such a defence of (P1') will be unaffected by MMNS’s methodological concerns.

There is a slight complication. As it stands, (P1') does not suffice for Churchland’s purposes. The issue is that (P1') does not specify which descriptions serve to fix the referents of the terms of folk psychology. Churchland assumes that the reference-fixing descriptions are the laws of folk psychology—so that, for example, the terms satisfy:

(6) For theoretical concepts of folk psychology $x_1, \ldots, x_n$: $x_1, \ldots, x_n$ only refer if some internal states $s_1, \ldots, s_n$ come sufficiently close to satisfying the folk psychological laws, in which case $x_i$ refers to $s_i$.\(^7\)

On this assumption, the radical falsity of folk psychology implies that there are no beliefs, desires, etc.\(^8\) However, one could endorse (P1') while denying (6). For example, one might claim that each concept of folk psychology is to be associated with a single description—perhaps ‘the state of holding a proposition true’ for the concept belief, and ‘the state of preferring a proposition to be true’ for desire, and so on. Call that single description descriptivism. Now, folk psychology could be radically false even if there are bona fide states of holding a proposition true and preferring a proposition to be true: it is at least a possibility that there are states of holding a proposition true, preferring a proposition to be true, etc., which do not come close to satisfying the laws of folk psychology. Thus, the assumption of a descriptivist theory of reference per se is not sufficient for Churchland’s argument from reference; the assumed theory of reference must be (6), or something very much like it.

---

\(^7\) I make the simplifying assumption throughout that, if there are states that come ‘sufficiently’ close to satisfying the laws of folk psychology, then those states come closest to satisfying the laws of folk psychology.

\(^8\) Assume (6). Now, suppose that there are beliefs, desires, etc. Then, by (6), beliefs, desires, etc., come ‘sufficiently’ close to satisfying the laws of folk psychology; and so the laws of folk psychology are not radically false; and so folk psychology is not radically false. By contraposition, we obtain: if folk psychology is radically false, then there are no beliefs, desires, etc.
As such, one should seek to respond to MMNS’s critique by replacing (P1) with the following premise:

(P1′′) The common-sense mental-state concepts are (to be explicated by the) theoretical concepts of folk psychology whose extensions are fixed descriptively by the laws of folk psychology.

One might then extend the original defence of (P1) to support the new (P1′′), by arguing for an explication of the common-sense concepts of belief, desire, etc., such that the explicata are the theoretical concepts of folk psychology construed so that (6) is satisfied.

As I construed Churchland’s defence of (P1), it aimed to justify explicating the common-sense mental-states concepts in accordance with (4). Here are five ways to precisify (4):

(7) (i) The common-sense mental-state concepts (such as belief, desire, etc.) are to be explicated; the explicata are the theoretical concepts of folk psychology, where it is stipulated that the extensions of the concepts of folk psychology are to be fixed descriptively, and it is stipulated that the reference-fixing descriptions are the laws of folk psychology.

(ii) The common-sense mental-state concepts (such as belief, desire, etc.) are to be explicated; the explicata are the theoretical concepts of folk psychology, where it is stipulated that the extensions of the concepts of folk psychology are to be fixed descriptively, and a single description is supplied for each concept.

(iii) The common-sense mental-state concepts (such as belief, desire, etc.) are to be explicated; the explicata are the theoretical concepts of folk psychology, where it is stipulated that the extensions of the concepts of folk psychology are to be fixed descriptively, but it is not stipulated what the reference-fixing descriptions are.

(iv) The common-sense mental-state concepts (such as belief, desire, etc.) are to be explicated; the explicata are the theoretical concepts of folk psychology, where it is stipulated that the extensions of the concepts of folk psychology are to be fixed causal-historically.

(v) The common-sense mental-state concepts (such as belief, desire, etc.) are to be explicated; the explicata are the theoretical concepts of folk psychology, where
it is not stipulated how the extensions of the concepts of folk psychology are to be fixed.

If one can justify (7i), then one can justify (P1′). Otherwise, not. In what follows I extend Churchland’s arguments, yielding arguments that (7i) is a better explication than (7ii–v).

To begin, then, recall (I): an explicatum should be precise. I submit that one way of making an explicatum precise is to stipulate how its extension is fixed. For example, consider the case of the concept planet. Suppose that the IAU had the following two candidate explicata:

\[
\begin{align*}
(8) \quad & (i) \quad \text{The concept } \text{planet}_1 \text{ is defined by } \varphi, \text{ or has sense } \varphi, \text{ or has stereotype } \varphi, \text{ etc.,} \\
& \text{and the extension of } \text{planet}_1 \text{ is fixed descriptively as whatever satisfies, or comes sufficiently close to satisfying, } \varphi. \\
& (ii) \quad \text{The concept } \text{planet}_2 \text{ is defined by } \varphi, \text{ or has sense } \varphi, \text{ or has stereotype } \varphi, \text{ etc.,} \\
& \text{but it is not stipulated precisely how the extension of } \text{planet}_2 \text{ is fixed.}
\end{align*}
\]

where “\(\varphi\)” abbreviates the tripartite condition developed by the IAU, viz. orbiting a star but not a planet, having an appropriate size, and having cleared its orbit. The concept \(\text{planet}_1\) is more precise than the concept \(\text{planet}_2\) in the sense that, relative to some (epistemically possible) state of the world, the extension \(\text{planet}_1\), but not of \(\text{planet}_2\), is fixed. When we use “\(\text{planet}_1\)”, we talk about those objects that satisfy \(\varphi\); when we use “\(\text{planet}_2\)”, it is unclear whether we are talking about (say) the objects that satisfy \(\varphi\), the objects that satisfy some other description, or the objects that are conspecific with those that are typically called “\(\text{planet}_2\)”, or something else. All else being equal, an explicatum is more precise if one has stipulated how its extension is fixed.

Now, (7i), (7ii) and (7iv) stipulate how the extensions of the explicata are fixed; but (7iii) and (7v) do not. In particular, (7iii) remains neutral on which descriptions fix the extensions, and (7v) remains neutral on what kind of mechanism fixes the extensions. In light of (I), then, (7i), (7ii) (7iv) are better explications than (7iii) and (7v).

So which of (7i), (7ii) and (7iv) is the better explication? Recall the earlier discussions of explanation and prediction of behaviour and of the problem of other minds. I will now argue that those two discussions can be extended to support the claim that (7i) is the better explication.

First, recall the discussion of explanation and prediction of behaviour. I presented an argument that the theoretical concepts of folk psychology are similar to our common-sense mental-state concepts in the following important respect: it is a distinctive feature of the
common-sense mental-state concepts that they are deployed in everyday explanations and predictions of behaviour, and the theoretical concepts of folk psychology feature in laws that underpin comparable explanations and predictions of behaviour.

Now, this distinctive feature of the common-sense mental-state concepts is preserved by (7i) but not by (7ii) or (7iv). Suppose that we treat the terms and concepts of folk psychology as though their extensions are fixed descriptively by the laws of folk psychology. Then, any competent user of those terms and concepts will grasp the laws of folk psychology, and will be able to use those terms and concepts to connect up the external conditions with internal states on the one hand, and internal states with an agent’s actions on the other. Such a user would thus be able to explain and predict behaviour in the ordinary manner.

Suppose instead that we treat the terms and concepts of folk psychology as either: having extensions fixed by single descriptions; or having extensions fixed causal-historically. Then, competence with those terms and concepts does not require one to grasp the laws of folk psychology, but rather to grasp either single descriptions for each concept or to stand at the end of appropriate causal-historical chains. As such, a competent user of those terms might neither be able to connect up internal states with external conditions on the one hand, nor be able to connect up internal states with an agent’s behaviour on the other. To explain how such users explain and predict behaviour in the ordinary manner, we would have to attribute to them an independent belief in the laws of folk psychology.

These considerations provide support for explicating the common-sense mental-state concepts in accordance with (7i) rather than (7ii) or (7iv). Consider (II): explicata should be similar in relevant respects to the explicanda. For Churchland, a distinctive feature of the explicanda—the common-sense mental-state concepts—is that they are deployed in everyday explanations and predictions of behaviour. The explicata—the theoretical concepts of folk psychology—themselves support comparable explanation and prediction of behaviour, only if they are construed descriptively with the laws of folk psychology acting as the reference-fixing descriptions. If the explicata are construed in terms of single description descriptivism or the causal-historical theory of reference, then the explicata would require supplementation with the laws of folk psychology in order to support the relevant explanations and predictions. So, given (7i), the explicata per se capture this distinctive feature of the explicanda (viz. that they are deployed in everyday explanations and predictions of behaviour). But, given (7ii) and (7iv), the explicata per se do not. In this respect, then, (II) supports (7i) over (7ii) and (7iv).

In addition, I note that criterion (III)—that explicata should be fruitful concepts—provides support for (7i). We have just seen that the concepts of folk psychology without
supplementation support the explanation and prediction of behaviour only when they are construed descriptively with the laws of folk psychology acting as the reference-fixing descriptions. Thus, in at least this respect, the concepts of folk psychology are more fruitful if they are so construed, rather than in terms of single description descriptivism or the causal-historical theory of reference.

Second, turn to Churchland’s discussion of the problem of other minds. The challenge, recall, was to explain how knowledge of other minds is in principle possible without appealing to ‘deduction from behaviour’ or ‘induction from the perilously isolated instance of one’s own case’. From the perspective of folk psychology, this challenge appeared to be easily met: ascriptions of mental states are hypotheses that would in principle be confirmed, were the explanations and predictions supported by those ascriptions sufficiently successful. Thus, (5), reprinted here, is an important consequence of this approach.

(5) In principle, our everyday explanatory and predictive practices could lead to knowledge of others’ minds (i.e. knowledge that another agent believes, desires, etc.).

However, this response to the problem of other minds relies on the assumption that the theoretical concepts of folk psychology are construed descriptively, with the laws of folk psychology acting as the reference-fixing descriptions.

First, suppose that the concepts of folk psychology are indeed construed descriptively with the laws of folk psychology acting as reference-fixing descriptions. On such a view, the ascription of mental states is simply the hypothesis that the agent has some internal states that (to a sufficient) degree satisfy the laws of folk psychology. Now, in principle, this hypothesis could be confirmed by our everyday explanatory and predictive practices: if those explanations and predictions are sufficiently successful—something which Churchland denies—, then one would be entitled to infer that there are indeed internal states that satisfy the laws that underpin those explanations and predictions. Thus, in principle, one could come to know that an agent is subject to states that satisfy the laws of folk psychology. And, given the supposition that the laws of folk psychology are reference-fixing descriptions, one could thus come to have knowledge that others have beliefs, desires, etc.

Suppose now that the concepts of folk psychology are construed descriptively, in terms of single description descriptivism. Then (5) does not follow from Churchland’s account of the problem of other minds. The success of our everyday explanations and predictions of behaviour can only confirm that there are some (nonspecific) states that come sufficiently close to
satisfying the underlying laws, not that the particular states of ‘holding a proposition true’, ‘preferring a proposition to be true’, etc., come sufficiently close to satisfying those laws. That is, assuming single description descriptivism, it is a possibility that: the laws of folk psychology are radically false because the states of ‘holding a proposition true’, ‘preferring a proposition to be true’, etc., do not come close to satisfying the laws of folk psychology; but, nonetheless, the explanations and predictions underpinned by folk psychology are highly successful because there are other internal states of agents that do come close to satisfying the laws of folk psychology. But then, our everyday explanatory and predictive practices would not lead to knowledge agents have beliefs, desires, etc.: however successful those practices, one would not be entitled to infer that an agent was subject to the familiar mental states.

The problem carries over mutatis mutandis on the supposition that the concepts of folk psychology are construed causal-historically: the success of our everyday explanatory and predictive practices could only lead to knowledge that some states come sufficiently close to satisfying the laws of folk psychology, not that the particular states that stand at the beginning of the relevant causal-historical chains come close to satisfying the laws of folk psychology. As such, however successful our explanatory and predictive practices, one would not be entitled to conclude that an agent has beliefs, desires, etc.

Churchland’s treatment of the problem of other minds, then, relies on construing the concepts of folk psychology descriptively, with the laws of folk psychology acting as reference-fixing descriptions. In at least this respect, the concepts of folk psychology are more fruitful if they are so construed, rather than in terms of single description descriptivism or the causal-historical theory of reference. Given (III), this is additional support for explicating the common-sense mental-state concepts in accordance with (7i), rather than in accordance with (7ii) or (7iv).

If the foregoing is right, then there are arguments available that support explicating the common-sense mental-state concepts descriptively, with the laws of folk psychology acting as the reference-fixing descriptions. But, then, (P1″) follows trivially. If one is within a theoretical context, then, given (7i), one should treat the terms of the familiar mentalistic vocabulary as theoretical terms of folk psychology construed descriptively. The premise (P1″) can be understood so as to merely make this commitment explicit.

This alternative formulation of Churchland’s argument avoids MMNS’s critique. The assumption of a theory of reference (which is explicit in (P1″)) is justified by an explication of the familiar mental-state concepts. There are independent arguments available to Churchland that the explication is, by the lights of (I), (II), (III), a good explication. However, as
explications are stipulative, they are not affected by concerns about the methodology involved in theorising about reference. That is, the soundness of the alternative formulation of Churchland’s argument from reference depends neither upon whether there is cross-cultural variation in semantic intuitions nor upon whether such variation undermines the standard methodology for theorising about reference.

4. Two objections

Objection one. “The alternative argument from reference does not establish Churchland’s intended conclusion. Churchland is concerned to show that it is a serious possibility that there are no beliefs, desires, etc., in the ordinary sense of those terms. The alternative argument from reference, in making use of explicata rather than explicanda, cannot establish this conclusion.”

Reply one. The objection misses its mark. My aim has been to provide an interpretation of Churchland’s argument from reference that illustrates how explications can be used in defence of arguments from reference, rather than to defend Churchland per se. Nonetheless, there is an important objection in this area, as follows: “what is really interesting about this issue is whether there are beliefs, desires, etc., in the ordinary sense of those terms. The alternative argument from reference, in making use of explicata rather than explicanda, does not speak to this issue.”

Let me briefly sketch my response to this objection. It is far from clear that, with their ordinary senses, the concepts of belief, desire, etc., are appropriate for serious theoretical inquiry into the mental architecture of humans. In any field of research in which linguistic precision is required, explication (or something like it) is required: our ordinary discourse is not sufficiently precise, unified or widespread to facilitate careful theorising. For example, it is not appropriate for cosmologists to discuss planets ‘in the ordinary sense of the term’; nor for physicists to discuss mass ‘in the ordinary sense of the term’; nor for economists to talk about markets ‘in the ordinary sense of the term’; nor for mathematicians to talk about infinity ‘in the ordinary sense of the term’; etc. I think that we should be highly sceptical of the claim that it is appropriate to discuss human mental architecture by talking about beliefs, desires, etc., ‘in the ordinary sense of the terms’. Such a discussion is only likely to be informative and substantial once we have clarified—i.e. explicated—the terms.

The underlying idea here is this: if one is interested in making substantial claims about human mental architecture, then one should not seek to ask whether there are beliefs, desires, etc., in the ordinary sense of the terms. Rather, one should seek to ask whether there are beliefs,
desires, etc. The appropriate tools for making such substantial claims are the explicata, not the explicanda.

It is worth noting that this point would not be that controversial in other domains. For example, suppose that a speaker, in an ordinary conversation, intends to make a substantial cosmological claim. Then it would be appropriate for her to use the explicated sense of “planet”; substantial cosmological claims are not held ransom by speakers’ intuitions about (say) “planet” or “Pluto”. Thus, she might appropriately use “Pluto is not a planet” to assert that Pluto is not a planet. Alternatively, suppose that another speaker, in an ordinary conversation, intends to make a substantial mathematical claim. Then it would be appropriate for her to use the mathematic sense of “infinity”; it would be appropriate for her to use “some infinite sets are larger than others” to claim that some infinite sets are larger than others. It would be irrelevant if one provided evidence that ordinary English speakers intuit, on the basis of their understanding of the words involved, that “some infinite sets are larger than others” is false. When engaging in a subject matter, one should use appropriate tools. And when making substantial claims about human psychology, one should use the appropriate, theoretical terms that one has available—such as the explicated, theoretical concepts of belief, desire, etc.

Objection two.9 “It is illegitimate to appeal to an explication of common-sense mental-state concepts in order to argue that there are no beliefs, desires, etc. The conclusion undermines the initial explication: empty concepts do not make good explicata. So the alternative argument fails: it is illegitimate to appeal (7i) in order to argue for eliminative materialism.”

Reply two. For ease of discussion, let us use “descriptive theoretical concepts” to denote the theoretical concepts of folk psychology construed so that their extensions are fixed descriptively by the laws of folk psychology. Thus, (7i) states that the common-sense mental-state concepts are to be explicated by the descriptive theoretical concepts. Now, independently of any explications, the alternative argument from reference I have provided is committed to the claim that the descriptive theoretical concepts are empty. (This is clear given that explicating the common-sense mental-state concepts with the descriptive theoretical concepts supposedly leads to the conclusion that there are no beliefs, desires, etc.) The objection is that the alleged empltness of the descriptive theoretical concepts undermines (7i).

Now, it is right that, when deciding between explications, the alleged emptiness of the descriptive theoretical concepts should be taken into account. But, contrary to the objection, I

---

9 Thanks to an anonymous referee for this volume for raising this concern.
do not think that the alleged emptiness automatically undermines (7i). The question is whether, given the emptiness of the descriptive theoretical concepts, the criteria (I)–(III) favour a different explication of the common-sense mental-state concepts whose explicata are non-empty.\footnote{An alternative interpretation of the objection would be that Carnap’s criteria should be amended, in order to rule out empty explicata. As I am assuming Carnap’s criteria, I will not consider this version of the objection here.} If there is no such explication, then the conclusion that there are no beliefs, desires, etc., stands. However, if there is such an explication, then the appropriate conclusion of the argument from reference should instead be that there are beliefs, desires, etc.

There are two points to make about this. First, either way, the principal aim of the paper is unaffected. In light of the above paragraph, we might ask: the assumption of which theory of reference can be justified by explication for deployment in the argument from reference? Depending on the answer to the question, the appropriate conclusion will either be that beliefs, desires, etc., do exist or that they do not exist. But, either way, this concedes the principal point that the assumption of a theory of reference in an argument from reference can in principle be justified by an explication.

Second, it does not suffice as an objection simply to note that there might be an alternative explication of the common-sense mental-state concepts. The objector would have to argue that there are alternative, non-empty explicata for the common-sense mental-state concepts, which better satisfy (I)–(III) than the descriptive theoretical concepts. And anyone who accepts the initial defence of (7i) is unlikely to accept that there are such alternative explicata.

For example, Churchland (1981: 72–76) offers a number of considerations in favour of the conclusion that nothing comes close to satisfying the laws of folk psychology. And, if that conclusion is right, then any non-empty explicata will either fail to satisfy criterion (I), which demands similarity to explicanda, or criterion (III), which demands fruitfulness. Suppose, first, that the non-empty explicata feature in generalisations that underpin explanations of human behaviour. Then, for Churchland, those explanations must be quite unlike the everyday explanations in which the common-sense mental-state concepts feature—otherwise we would expect the common-sense mental-state concepts to likewise be non-empty. As such, for Churchland, the non-empty explicata would in this key sense be dissimilar to our common-sense mental-state concepts and would thus fail to satisfy (I). But if the non-empty explicata did not feature in any generalisations that underpin explanations of human behaviour, then they would not be fruitful in Carnap’s sense; they would not satisfy criterion (III). Either way, from
Churchland’s perspective, it is unlikely that there are non-empty explicata for the common-sense mental-state concepts that better satisfy (I)–(III) than the descriptive theoretical concepts.

5. Closing remarks

An argument from reference is an argument that assumes a theory of reference and seeks to establish a substantial philosophical conclusion. MMNS critique arguments from reference in three stages. First, they present evidence that there is cross-cultural variation in semantic intuitions. Second, they argue that such cross-cultural variations undermines the standard methodology for theorising about reference. Third, they argue that, without an acceptable methodology for theorising about reference, the assumption of a theory of reference in any argument from reference is unjustified. Extant responses to the critique only discuss the first and second stages. However, for proponents of arguments from reference, this is insufficient. In particular, with respect to Churchland, neither cross-cultural variation in folk semantic intuitions, nor the consequence for methodology, should be relevant to whether or not there are beliefs, desires, etc.

I have looked at Churchland’s argument from reference in detail, providing an interpretation of that argument from which I have drawn out and motivated a response to the third stage of the critique: one can justify the assumption of an appropriate theory of reference by appealing to an explication of the relevant common-sense concepts. As explications are stipulative, they are independent of MMNS’s methodological concerns for theorising about reference.

Although I have not argued explicitly for this conclusion here, I take the preceding discussion to exemplify a general strategy for responding to MMNS’s critique. Whenever the assumption of a theory of reference in an argument from reference can be justified by appeal to an explication, the third stage of MMNS’s critique fails. The strategy may not be successful for every extant argument from reference; but I see no reason to suppose it can only be successful for Churchland.11 It is unclear whether we can say anything general about when this strategy for responding to MMNS’s critique will be successful. But this is not surprising. Whether an explication is a good explication will depend on the specifics of the case.

11 For example, the strategy can be straightforwardly used to support Andreasen’s (2000) argument that race is real.
Earlier, I adopted Carnap’s three principal criteria, (I)–(III). Note, however, that these criteria are not essential to the general strategy. The key point is that explications are stipulative, and thus avoid MMNS’s methodological concerns, not that explications satisfy (I)–(III). So, for example, one might amend some of the existing criteria (see notes 4 and 10), change how they are weighted (see e.g. Kitcher 2008), or add additional criteria (such as Carnap’s subordinate criterion of simplicity (1950: 7)). So long as there are appropriate ways to stipulate a refinement of a pretheoretic concept for theoretical purposes, one may sometimes stipulate that the refined concept be construed in accordance with one or another theory of reference; and when one can do that, one is justified in assuming a theory of reference in an argument from reference.

And that is so regardless of whether there is cross-cultural variation in semantic intuitions, and regardless of whether such cross-cultural variation would undermine the standard methodology for theorising about reference.12

References


12 For helpful comments and discussion of drafts of this paper and/or the ideas contained within, I would like to thank James Andow, Emma Borg, Michael Devitt, Anthony Everett, Jumbly Grindrod, Nat Hansen, Krzysztof Poslajko, Kathy Puddifoot, anonymous referees for this collection and a journal, and audiences at PhiLang 2015, the Bucharest Colloquium in Analytic Philosophy 2015 and the University of Reading. I gratefully acknowledge the financial support of the Analysis Trust, who have funded this research.


