Cogency and Question-Begging: Some Reflections on McKinsey’s Paradox and Putnam’s Proof

Crispin Wright

I

Preliminaries

A cogent argument is one whereby someone could be moved to rational conviction of the truth of its conclusion. So a chain of valid inferences cannot be cogent if only someone who already took themselves to be rationally persuaded of the conclusion could rationally receive whatever grounds purportedly warranted its premises as doing just that.

Say that a particular warrant, \( w \), transmits across a valid argument just in case the argument is cogent when \( w \) is the warrant for its premises. We need to note immediately a distinction between transmission of warrant, so characterised, and closure of warrant. Closure of warrant across entailment has of course been very widely discussed.\(^1\) It is the weaker principle. Closure, unrestricted, says that whenever there is warrant for the premises of valid argument, there is warrant for the conclusion too. Transmis-
sion, unrestricted, says more: roughly, that to acquire a warrant for the premises of a valid argument and to recognise its validity is to acquire—perhaps for the first time—a warrant to accept the conclusion. My concern in this note will be with a particular type of counterexample to transmission. Such cases need not be counterexamples to closure. Closure would hold, for instance, but transmission might fail, in a case where there was warrant for the premises in the first place only because the conclusion was antecedently warranted. (I am, myself, sceptical whether there are any genuine counterexamples to closure but that issue is not on our agenda.)

Transmission of warrant need not be an absolute characteristic of a valid argument. It may be that a particular argument is such that one type of possible ground, \( w_1 \), for its premises is transmissible—can yield a novel reason for accepting the conclusion when taken in conjunction with recognition of the validity of the inference—while another, \( w_2 \), is not, but can only be possessed in the first place by a thinker whose information already includes warrant to accept the conclusion.

Intuitively, a transmissible warrant should make for the possible advancement of knowledge, or warranted belief. A warrant is transmissible, more specifically, when we may envisage a logically non-omniscient but otherwise perfectly rational subject coming to believe a proposition for the first time in a way which depends on their recognition both of the validity of the inference in question and of their possession of warrant for its premises. So there is one immediate and very simple kind of counterexample to transmission which is not a counterexample to closure: that of simple question-begging—the case where the conclusion of a valid argument features explicitly among its premises. In that case, recognition of the validity of the inference will, obviously, have no part to play in a subject’s recognition of their warrant for its stated conclusion. But there are more interesting ways of begging questions than that.

II

Transmission failure for information-dependent warrants

Here is one kind of example. It comes by way of a by-product of the “theory-ladenness” of empirical confirmation. You bring a ket-
tle to the boil on a camping stove: is that evidence that the li-
uid inside is at or close to 100° Celsius? Yes, if you have it that
the liquid is water and the atmospheric pressure is around that of
sea level; no, if you know the liquid isn’t water or if you are
making tea at Everest base camp. You see a soccer game appar-
ently in progress: one of the players drives the ball into the net
and turns to receive the congratulations of his colleagues while
the crowd cheer and the referee turns and points to the centre
circle. Is all that evidence that a goal has been scored? Yes, in
normal circumstances; but no, if you happen to know that this is
the third take of a scene for a television drama. The familiar
point these examples make is that in a wide class of cases, the
fact of whether particular considerations provide a warrant for a
particular belief depends on the character of one’s collateral
information.

Let’s pursue the soccer example in the light of that point. Con-
sider the inference from

(i) A goal has just been scored,

to

(ii) A game of soccer is in progress.

Assuming—as we may—that it is only in the context of a genuine
game of soccer that there is such a thing as scoring a soccer goal,
this is a valid inference. Let the evidence for (i) be as described—
the driving of the ball into the net, the apparent congratulations
of team-mates, the referee’s response, etc. Is any warrant supplied
by this evidence for (i) transmissible across the entailment to (ii)?

Well, not at least in the following scenario. Suppose you are in
the vicinity of a film studio, and know that it is just as likely
that the witnessed scene is specially staged for the camera as
that it is an event in a genuine game. Once you’re equipped with
this information, you will rightly regard the described evidence as
providing no warrant any longer for (i). What you need, in order
to restore its warranting power, is precisely some independent cor-
roboration of (ii). You ask a bystander: is that a genuine game or
a film take? When you learn the game is genuine, you once again
have a warrant for the claim that a goal was just scored. But it
would be absurd to regard that warrant as transmissible across
the entailment to (ii). You don’t get any additional reason for
thinking that a game is in process by having the warrant for (i).
It remains that your only ground for (ii) is the bystander’s testi-
mony and it is only because you have that ground that witnessing the scene provides a warrant for (i). So that's a very clear case of transmission failure. (And note, interestingly, that there is no of failure of closure: it remains true in the described scenario that if one does have a warrant for (i), one also has a warrant for (ii).)

It would be easy to multiply further examples of this kind. The key point is the pervasive holism of confirmation—or as I prefer, the information-dependence of warrant. A body of evidence, e, is an information-dependent warrant for a particular proposition P if regarding e as warranting P rationally requires certain kinds of collateral information, I. Some examples of such e, P and I—like those in the soccer example—have the feature that elements of the relevant I are themselves entailed by P (together perhaps with other warranted premises). In that case, any warrant supplied by e for P will not be transmissible to those elements of I. Warrant is transmissible only if a rational thinker could cite among her reasons for accepting I the fact that she has warrant for P together with the entailment. But at least in the described scenario, it would be absurd for a thinker to cite e's warranting P as a reason for accepting I, since it is only on the assumption that I is independently warranted—by the bystander's testimony—that she has any warrant for P in the first place. Her warrant for P is thus not to be reckoned as among her reasons for accepting I—either as bestowing a first such reason, or as enhancing reasons already possessed.

III

The paradox

Michael McKinsey\(^2\) is responsible for an influential presentation of a kind of argument—actually, a paradox—which purports to elicit an inconsistency between (allegedly) plausible externalist constraints on content and what is often termed first-person privileged access—the combination of groundlessness and authority standardly possessed by a subject's opinions about her own intentional and other psychological states. This kind of argument may be represented extremely simply in the following (MC) form:\(^3\)

\[
\begin{align*}
\text{I have mental property M;} \\
\text{If I have mental property M, then I meet condition C.}
\end{align*}
\]
Therefore

I meet condition C.

An example would be

(1) I believe that water is wet
(2) Any thinker who believes that water is wet (belongs to a speech community which) has had such-and-such encounters with water

Therefore

(3) I (or my community) have had such-and-such encounters with water.

What's the paradox? Well, assume that reflection on the nature of content in general and the character of the concept, water, in particular, enables us to recognise plausible externalist constraints in virtue of which the second premise holds, and hence that it can be warranted a priori, by pure philosophy. If, as our ordinary conception of the psychological allows, it is possible for me warrantedly to accept the first premise without reliance on any kind of empirical evidence or other articulated grounds, then it appears that a warrant can be achieved for both premises purely by reflection, without empirical enquiry, and hence—if transmission holds here—that such a purely reflective warrant may consequently be acquired for the conclusion, just by adding in a (reflective) recognition of the entailment. Yet the idea that I can come to a warranted belief that my (community's) history includes interaction with water, or any other particular stuff, purely reflectively, without reliance on any kind of empirical investigation, seems quite preposterous.

Of course the problem might well be felt to be with the second, externalist premise. I am myself by no means certain that there is any compelling form of content externalism which will sustain a paradox-generating version of that premise. Reflect, for instance, that the stereotype of a natural kind concept which the literature has gathered, rightly or wrongly, from the writings of Putnam and Kripke would seem to be one for which possession of such a concept is compatible with reference failure. Water, conceived as a natural kind concept à la Putnam/Kripke, picks out, if anything, the actual kind of stuff whose having the physical essence it has underlies and explains the characteristic ways in which what we take to be water presents itself. When Twin Earth is as in Putnam's story, that concept does not include Twin Earth's lakes.
and rivers in its extension. Clearly we could set up a concept with the intention that its extension be determined in the indicated kind of way and then be frustrated—if there were, as it happened, no essential characteristics in common among (enough) typical manifestors of watery symptoms. But it would be hard to deny that we had so much as fixed a concept: the concept is precisely that of the kind of stuff whose having the physical essence... etc. If the strongest plausible form of externalism about natural kinds is one compatible with this broad conception of a natural kind concept, then no troublesome version of the second premise is going to hold good a priori.

Externalism about this kind of concept may involve all or only some of a variety of claims: for instance, that the concept is rigidly associated with its actual extension, that it is identified by its extension, that it cannot so much as exist without a (non-empty) extension. The last is what is germane to the second premise of the McKinsey paradox. But the question, just which of these claims are really well motivated by the basic externalist arguments and ‘intuitions’, is delicate and unresolved, so it seems to me. And explicit argument for the ‘no (non-empty) extension—no existence’ claim is rare.5

However I propose to finesse this issue. My interest in what follows is entirely in the status of the McKinsey argument if there is indeed adequate motivation for a kind of strong externalism sufficient to drive its second premise.

IV

Davies’ proposal

Martin Davies has recently advanced the arresting suggestion that, whatever else might be amiss with them, McKinsey-style arguments may provide an example of transmission failure.6 Certainly, the paradox does wholly revolve around the assumption of transmission of warrant: specifically, around the thought that to concede that the premises in a valid argument are warranted purely by reflection entails that the conclusion may be so warranted as well. But Davies’ suggestion may seem to come right out of the blue nonetheless. Why suspect transmission at all in such cases?

Davies does not explicitly address that question (except by his developed account) but a prima facie motive may be provided by
the following analogy. Russell canvassed the sceptical possibility
that the world sprang into existence five minutes ago replete with
apparent traces of a much more ancient history. Suppose an ar-
chaeologist accumulates orthodox evidence for the impact of a
large meteor on the earth's surface some 140 million years ago—he
finds a characteristic crater formation, with chemical and other
changes in the local rocks consistent with a high energy impact;
investigation of other nearby fragments of rock finds them to be
exceptional in the context of the local geology, etc. Clearly we
now have the materials for a set-up to which the transmission-
limiting considerations outlined above, connected with informa-
tion-dependence of warrant, would apply directly. That is, our
archaeologist has good but defeasible evidence for the claim

(i) A meteor impacted on the earth's surface in this region some
140 million years ago,

which in turn entails denial of Russell's Possibility. However, the
character of the evidence for (i) is such that there is no question
of it extending to the exclusion of Russell's Possibility in turn,
since it is only if one already excludes that possibility that the
type of evidence mentioned is supportive of (i) in the first place.

Now, a rejection of Russell's Possibility is, of course, also en-
tailed by the conclusion of the McKinsey argument: if I or my
speech community has indeed had the appropriate kind of encoun-
ters with water, then it cannot be true that the earth came into
being just five minutes ago, replete with traces of an apparently
much more ancient history. Accordingly that argument readily ex-
tends itself into one whose conclusion is likewise that Russell's
Possibility does not obtain. And if warrant is indeed transmissi-
able from its premises to the conclusion that my life history in-
cludes certain kinds of encounter with water, there would seem to
be nothing to stop the warrant extending to the denial of Rus-
sell's Possibility as well.

Both arguments, then, may terminate in the denial of what Rus-
sell took to be an intractable metaphysical possibility. Both are
valid. Both proceed from premises which one would normally sup-
pose to be entirely warrantable. Both seem to prove far too much.
How might the fact that the one works with premises which are non-empirically warranted secure for its conclusion a cogency
which the other—working with routine historical evidence—cannot
provide?
That’s enough to at least motivate interest in the thought that there might be a transmission problem in the McKinsey case. Exactly the difference just emphasised, however, ensures that anyone inclined to think that there is something suspect about transmission in both cases will have to find rather different explanations of what and why. For, to stress, there is, in the archaeological case, a separately characterisable body of evidence whose ability to provide a warrant for the proposition about the meteor needs the setting of the prior assumption that Russell’s Possibility does not obtain. What is striking about the McKinsey case, by contrast, is that evidence plays no role in the warrant for either of its premises. Rather, the first premise is precisely groundless, after the fashion of many routine psychological self-ascriptions; and the second is, supposedly, warranted a priori, by reflection. Since there is, in the relevant sense, no empirical evidence in either case, there can be, it would seem, no information-dependence of evidential force—no collateral beliefs on whose satisfaction the status of the evidence might depend.

Davies is quite clear about this and responds by proposing a new limitative principle:

\[(D) \text{Warrant cannot be transmitted across a known entailment if the truth of the consequent is a precondition of the subject’s capacity to believe the antecedent.}\]

Since the second-order belief—my belief that I believe that water is wet—involves an exercise of the concept water and thus—in the presence of the externalist premise—makes exactly the same demands on my previous history as the first-order belief, I cannot believe the premise for the McKinsey argument—that I believe that water is wet—unless the conclusion is satisfied. So (D) will block the transfer of warrant in this case.

What is there, though, to motivate such a principle? Here is what Davies suggests:

The intuitive idea behind [principle (D)] is something like this. In any given epistemic project, some propositions will have a presuppositional status. Suppose that the focus of the project P is the proposition A,... then within P it is presupposed, for example, that A is a hypothesis that can be coherently entertained (can be believed, doubted, confirmed, disconfirmed);... suppose that B is some proposition that has this presuppositional status in project P. Then P cannot itself yield knowledge that B; nor can P play an essential role in yielding knowledge that B.
Davies’ idea, then, is that in any given ‘epistemic project’—a project in which we set ourselves to accumulate evidence on a certain question, or test a certain hypothesis, etc.—there will be certain things we have to take for granted. We will, for instance, take for granted a certain body of prior information and this, as stressed above, will then condition what may be counted as evidence for or against the propositions that are central in the project. But we will also inevitably take for granted whatever conceptual apparatus conditions the intelligibility of the project, and in particular we will take it for granted that we can so much as entertain such propositions as are essential to it. Davies’ intuition, in short, is that epistemic projects have presuppositions in a sense which involves that one thing they cannot deliver, under any circumstances, is (an enhancement of) warrant for their own presuppositions. One type of presupposition will be the collateral information which conditions the warranting force of the evidence which the project throws up. So the type of limitative principle operative in the soccer example is covered by Davies’ intuitive idea. But if he is right, then the relevant notion of presupposition—i.e., one which has the twin effect (a) that presuppositions are exactly things which epistemic projects have to have, and (b) which they cannot themselves confirm—is of wider application, sustaining both the principle (D), and perhaps further limitative principles as well.

Let’s call the general principle being mooted here—the suggestion that:

Every epistemic project incorporates certain presuppositions which, whatever its upshot, it cannot provide warrant for—

the PREP principle (presuppositionality of epistemic projects). It is an attractive principle in general terms—indeed, it would seem to be a core idea of Wittgenstein’s notes On Certainty. But, as Davies emphasises, it is of limited use to us until we are clear what the relevant notion of presupposition is—what notion of presupposition, that is, makes PREP true—and what things we should reckon among the presuppositions in that sense of any particular epistemic project. Principle (D) may seem to incorporate one not implausible assumption about that. But Davies is uneasy with it; and so—essentially for the reasons he gives—am I.

The main concern is that D is open to simple counterexamples—at least if “precondition” means: necessary condition—and that it is by no means clear how better to interpret the
17. COGENCY AND QUESTION-BEGGING

notion. It is for example, as Davies points out, a necessary condition of my ability to believe that I believe that water is wet that someone is capable of believing something. Yet it’s not evident that there is any reason to restrict the transmission of warrant in that example. If you were challenged to produce a reason for thinking that someone was capable of believing something—bizarre as the challenge would doubtless seem—it wouldn’t seem question-begging to point out that so much is entailed by your believing that water is wet.9 To be sure, one might have thought it evident from the start that making a plausible general case for PREP, or derivatives like D, would indeed depend on finding some narrow and more interesting notion of ‘presupposition’ than that. But all the work remains to be done;10 and if a doubt about transmission in McKinsey cases is to turn on its outcome, then that diagnosis of what is wrong with such arguments had better be provisional.

I think Davies’ general vision of the issues, incorporated in the PREP principle, is well worth further study. But in what follows I will restrict attention to trying to describe a rather more specific possible basis for the idea that transmission is the crux in the McKinsey argument

V

The connection with Putnam’s proof

The “Putnam’s proof” of my title is his celebrated argument that we are not brains-in-a-vat.11 What bearing has any of the foregoing on that argument? In earlier work,12 I proposed a formulation of the nerve of the “proof” as follows. Let BIVese be defined as the language of a brain-in-a-vat in Putnam’s scenario.13 Then

(i) In my language, any meaningful expression may be used disquotationally to specify its own reference—Premise.
(ii) In BIVese, “brain-in-a-vat” does not refer to brains-in-a-vat—Premise (backed by externalist constraints on reference and the details of the example).
(iii) In my language, “brain-in-a-vat” is a meaningful expression—Premise.
(iv) In my language, “brain-in-a-vat” refers to brains-in-a-vat—from (i) and (iii).
Hence
(v) My language is not BIVese—from (ii) and (iv).
But
(vi) If I am a brain-in-a-vat, my language, if any, is BIVese—
definition of BIVese.
So
(vii) I am not a brain-in-a-vat. QED.

The argument is clearly valid, and there is a case that, against
some popular objections, it is not merely valid but cogent. However
the precedent of the very simple McKinsey argument sud-
denly suggests that we may formulate the essential reasoning much
more tersely, without significant loss, in MC form:

(i) I have the concept of what it would be to be a brain-in-
a-vat (in Putnam’s scenario, etc.). (Self-knowledge premise)
(ii) No such brain-in-a-vat can have that concept. (External-
ist premise)
Hence
(iii) I am not a brain-in-a-vat.

That really does, it seems, get everything essential.
We might react by concluding that Putnam’s proof and Mc-
Kinsey’s argument are essentially the same!—and hence that,
though each is valid, Putnam’s can be no more—nor less—cogent
than McKinsey’s. So if we are about to find a transmission failure
in McKinsey’s argument, it will be there in Putnam’s proof too.
But the latter epitomises a whole tradition of transcendental arg-
uments that attempt to elicit metaphysical conclusions from prem-
ises about our understanding. Are they all to turn out to be
examples of transmission failure—if the McKinsey argument is
indeed such—or can we make a subtler discrimination? I shall
argue that we can.

VI

The co-warrantability of the McKinsey premises

There is a further natural reservation to have about the McKin-
sey argument even if one thinks some well-motivated form of con-
tent externalism is available to drive (something like) the second
premise: Can the premises of the argument be *simultaneously* yet non-empirically warranted in the fashion the paradox exploits?

Here is a line of argument which might lead one to doubt about that. Any version of externalism sufficiently strong to drive the second premise in the McKinsey argument has to be inconsistent with the idea that a natural kind concept, or other kind of concept to which it applies, might fail of reference altogether. At the same time, its being genuinely a variety of externalism requires that what determines the reference, hence identity of the concept in question, or indeed whether any reference is established at all, may lie outside the scope of the reflective knowledge of the subject. So if ‘water’ is to express, if anything, then a concept regulated by such an externalism, it is implicit in the case that a subject may think he has a concept, which he expresses by use of the term “water”, when in fact he has no concept at all. In other words, any form of externalism strong enough to drive the argument must be strong enough to set up the possibility of *illusions of content*. I may think I have a concept which I express by the term, ‘water’. But it may be clear to me that if I do indeed have such a concept, it is a natural kind concept in the sense that its very existence depends on its presenting a real property with whose instances I (or, depending on the story, others of my speech community) have interacted. But since that is a condition whose satisfaction I cannot certify merely by reflection of my own apparent mental states, it follows that I may be deluded in thinking I have any such concept at all.

There is a close parallel here with the semantics of (what he styles) *Russellian* proper names defended by Gareth Evans. Evans retains the sense/reference distinction for proper names generally. But in the case of Russellian names (which are, roughly, those fitted to subserve genuinely object-directed thought) it is only via the establishment of a referent that a proper name acquires a sense (so that it is wrong in such cases to think of sense in terms of Frege’s sometime proposal, as a *route* to reference, to which no actual referent may correspond). It goes with this view that a purportedly Russellian name for which the reference-fixing episode somehow aborted—for instance, if the whole thing occurred in an episode of communal hallucination—has no determinate sense and accordingly makes no contribution towards determining a truth-condition for sentences containing it. Yet a recipient of the abortive explanation may believe that she attaches a definite meaning to the term, and that there are determinate thoughts which she expresses by using it in sentences. That would be an illusion, as
Evans explicitly recognised. What I am suggesting, in parallel, is that any form of externalism strong enough to drive the McKinsey argument must make space for exactly comparable illusions, the only difference being that it is common nouns, not proper names, that are affected.

Suppose that's correct. Then doesn't it bring out—and just as one would naturally have suspected—that content externalism must after all compromise the epistemic sovereignty of First Persons over their attitudes and meanings, and so jeopardises the warrant for the first premise of the McKinsey argument?

Well, I think it's undeniable that there is a compromise, at least if the sovereignty was meant to involve a kind of infallibility about the content of one's attitudes—about what their content is, and when there is so much as a content there at all. It is less clear to me that a violation is involved of anything that should properly be regarded as belonging to our intuitive preconceptions about self-knowledge. Our intuitive preconceptions, it seems to me, insofar as they are definite, require no more than that the following three conditions be met:

(i) that a subject may normally be expected to know the content of her attitudes and what attitude she is taking to those particular contents,

(ii) that in a wide class of cases she will know of these matters non-inferentially, and

(iii) that ceteris paribus they are matters on which she should be treated as authoritative and to be deferred to.

Each of these three can survive the admission of the possibility of externally generated illusions of content. It is likely, no doubt, that ordinary thinking about the psychological fails to reckon with such a possibility, or in general with philosophically sophisticated considerations about mental content. But to fail to reckon with is not the same thing as to be inconsistent with.

Here is what I think is the crucial consideration for our present purpose—i.e., as far as the co-warrantability of the McKinsey premises is concerned: it is that the possibility of a certain kind of illusion of content should no more compromise our warrant for the ordinary run of self-ascriptions of content-bearing states than the possibility of perceptual illusions compromises our warrant for the ordinary run of claims about the immediately perceptible environment. Perceptual illusions do occur; but warrant for perceptual claims works in such a way that there is a default presumption
against them—our practice is that you don’t need to gather evidence that no illusion is operative in the circumstances before you are warranted in treating your experience of the cat on the mat before you as grounds for the claim that there is indeed such a cat, so situated. Rather you have a warrant until or unless grounds are forthcoming for thinking that an illusion is operative. There are obvious, pragmatic reasons why that has to be the order of things. So it should be with illusions of content. There is no reason why the mere possibility, if indeed demanded by any plausible form of externalism, should do anything to compromise our ordinary conception of the non-inferential evidence of a subject’s attitudes to her.

It therefore seems to me that the warrant for the first premise of the McKinsey argument is not compromised by the second—and the paradox is therefore with us still.

VII

Towards solving the paradox: why transmission fails

The point connecting any externalism strong enough to sustain the second premise with the possibility of content-illusion is nevertheless crucial, for it is that which sets up the worry about transmission in McKinsey cases. Here is the point in a small nutshell. Suppose that I want to describe myself truly as it happens by using the words: “I believe that water is wet”. If some form of externalism is true which is strong enough to sustain the second premise of the McKinsey argument, then there are external preconditions of my expressing a true belief by those words whose satisfaction I may nevertheless, without compromise of the warrant for my claim, have done nothing special to ensure. Can that warrant now licitly be extended to the claim that those preconditions are met? It should seem obvious that it cannot, for the simple reason that the warrant is in the first place conditional on the concession, as it were, that unless there is extant positive reason to doubt that the external conditions are met, the possibility that they are not can be ignored. Given that the warrant I start out with has this concession-dependent character, it naturally cannot be massaged by inference into a reason for a positive view about the issue which the concession was precisely a concession to take for granted.
That may seem a little terse, so consider some other examples—simpler, yet of the same essential structure—which make the same point. You go to the zoo, see several zebras in an enclosure, and opine that these animals are zebras. Well, you know what zebras look like, and these animals look just like that. Surely you are fully warranted in your belief. But if the animals are zebras, then it follows that they are not mules painstakingly and skilfully disguised as zebras. Does your warrant transmit across the entailment to the latter claim? There is a strong intuition that it does not. Did you examine the animals closely enough to detect such a fraud? Almost certainly not. The grounds you have for "These animals are zebras"—essentially, the look of the beasts—have no bearing on this possibility.

Again: you look at a wall and see that it is painted red. So you have acquired a warrant for thinking that it is red. But its being red entails that it is not a white wall cleverly illuminated by concealed lighting to look as if it is red. So have you thereby acquired a warrant for thinking that? Again, the strong intuition is not. Your warrant was acquired just by looking at the wall—no doubt you did enough to verify that it is red if indeed it is, but what you did simply didn’t reckon with the possibility of deceptive concealed illumination.

When, in the contexts described, you form your beliefs about the zebras and the colour of the wall, there are external preconditions for the effectiveness of your method—casual observation—whose satisfaction you will very likely, without compromise of the warrant you acquire for those beliefs, have done nothing special to ensure. Made-up mules and tricky lighting involve the frustration of those preconditions. Can the warrants you acquire licitly be transmitted to the claim that those preconditions are met—or at least that they are not frustrated in those specific respects? It should seem obvious that they cannot. While you have—no doubt quite properly—taken it for granted that the conditions were generally suitable for the acquisition of reliable information by purely perceptual means, it would be absurd to pretend that you had gained a reason for thinking so—at least in the specific respects that you didn’t have to reckon with disguised mules or deceptive lighting—just by dint of the fact that those specific possibilities are logically excluded by the beliefs which, courtesy of your background assumption, you have now confirmed.

To elaborate and generalise, consider this structure. Suppose I take myself to have a non-inferential warrant—perhaps perceptual, or introspective, or intuitive, or mnemonic—for some prop-
osition A. And let the question be whether this warrant transmits to a certain consequence, B, of A. However suppose A stands opposed to some proposition C which would be true in circumstances subjectively indistinguishable from those in which I actually find myself, and which does not entail B, but would be true if B were false. The set-up is thus

(i) that A entails B;
(ii) that my warrant for A consists in my being in a state which is subjectively indistinguishable from a state in which the relevant C would be true;
(iii) that C is incompatible with A; and
(iv) that C would be true if B were false.

The key question is what, in the circumstances, can justify me in accepting A? Why not just reserve judgement and stay with the more tentative disjunction, "A or C"?—for it is all the same which disjunct is true as far as what is subjectively apparent to me is concerned. The answer has to be, it would seem: because I am somehow additionally entitled to discount the other disjunct, C. It may be that I have collateral information telling against C. Or it may be that, in the sort of way mooted in the case of perceptual illusion, I am not required to bother about C. But either way, in order for me to be entitled to discount C, and so move past the disjunction to A, I have to be entitled to discount the negation of B, and therefore entitled to accept B; for by hypothesis, if not-B were true, so would C be. So it would seem that I must have an appreciable entitlement to affirm B already, independent of the recognition of its entailment by A, if I am to claim to be warranted in accepting A in the first place. So the inference from A to B is not at the service of a rational first conviction that B.

It is very straightforward to cast the zoo and red wall examples into this template. Let A be, respectively, the proposition that the animals in question are zebras or the proposition that the wall is red; B is accordingly the proposition that the animals are not mules painstakingly and skilfully disguised as zebras or the proposition that the wall is not a white wall cleverly illuminated by concealed lighting to look as if it is red; while for a suitable C we need look no further than the respective negations of those two choices for B. The experiences of seeing zebras and seeing mules painstakingly and skilfully disguised as zebras are relevantly subjectively indistinguishable; as are the experiences of seeing a red wall and seeing a white wall cleverly illuminated by
concealed lighting to look as if it is red. So in treating my visual experience as being of zebras or a red wall respectively, I implicitly discount the uncongenial, deceptive alternatives. And now, whatever my warrant for doing so, it has to be there already, independently of any consideration of the entailment of their respective negations by the appropriate choices for A.

How does it go with the McKinsey argument? Take A as the proposition that I believe that water is wet, B as the proposition that I, or my speech community, has had such-and-such encounters with water, and C as the proposition that the seeming-thought which I attempt to token by "I believe that water is wet" is content-defective owing the reference failure of the purported natural kind term, "water", in my language. Then each of the four conditions delineated is met:

(i) Proposition A, that I believe that water is wet, entails—on the assumption of the necessary truth of the relevant strong externalism—proposition B, that I, or my speech community, has had such and such encounters with water;

(ii) My warrant for A consists in my being in a state which is subjectively indistinguishable from a state in which the relevant proposition C, that the seeming-thought which I attempt to express by "I believe that water is wet" is content-defective owing the reference failure of the purported natural kind term, "water", in my language, would be true;

(iii) C is incompatible with A; and

(iv) C would be true if B were false.

And the consequential analysis of the transmission failure involved in the McKinsey argument—always assuming cogent a priori motivation for its externalist premise—is simply that once that premise is in play, my justification for discounting the uncongenial interpretation of my subjective state as one in which C holds—and hence for taking it that A is true—must depend on antecedent reason to think that my tokenings of "water" comply with appropriate externalist constraints, exactly as described by B. So in taking it that I am warranted in accepting A, I presuppose a warrant for B independent of the recognition of its entailment by A. The recognition of that entailment is thus not at the service of a rational first conviction of B.

Of course, if such an antecedent reason for discounting C had to be empirically acquired, this line of thought would after all involve dismissal of the groundlessness of self-knowledge in the
relevant kind of case; rather, evidence would be presupposed that "water" indeed possessed the appropriate historical connections. Even then, if the diagnosis is right, there would be a transmission failure. But I have suggested above that even if one effect of a correct externalism is indeed to introduce, via the possibility of illusions of content, a novel form of defeat for self-knowledge claims, the epistemological impact of that possibility ought to be conceived much as that of the possibility of perceptual illusion vis-à-vis the justification of perceptual claims: that is, plausibly, that we may reasonably allow an a priori presumption against it, since to allow an empirical issue to arise in every case would be to abrogate the means to resolve any such issues (I have to take certain perceptions to be reliable if I am to investigate the reliability of others; I have to take certain seeming-thoughts to be well-founded if I am to investigate the well-foundedness of others). If that is the right way to look at the matter, at least as a starting stance, then the groundlessness of self-knowledge—properly conceived—is unimpugned by externalism, we have a (defeasible) entitlement to set aside the uncongenial C without evidence, and the effect of the McKinsey argument is, not to make available a non-empirical warrant for its conclusion but to bring out the empirically unearned commitment to its conclusion, and to claims of that kind in general, on which first-person authority—for (sufficiently strong) externalism—rests.

It should be superfluous to remark, finally, than none of the examples, if here treated correctly, presents a counterexample to closure. On the contrary, it is built in to the diagnosis of the transmission failure involved that they do not. The argument has precisely been that a prior warrant to accept the relevant propositions B, appreciable independently of their entailment by the relevant propositions A, is a necessary condition for possession of the relevant kinds of warrant for the latter. So there will be warrant for the conclusions of the relevant arguments whenever there is (that kind of) warrant for their premises.

VIII

A contrast with Putnam’s proof

Further work would be wanted on the specific template proposed if we are to be confident that it is precisely right; it is not meant, of course, to capture all cases of transmission failure but it does
need to be seen that it excludes no intuitively bona fide cases of transmission. I shall not pursue the issue further on this occasion. If something close to the proposed template is sound, we may expect to conclude that whatever one should think about the major, externalist premise of the McKinsey argument, there is a mistake involved in the assumption of transmissibility which underpins its paradoxical aspect—the aspect which it wears if canvassed not merely as a valid argument but as a proof. It remains to see whether the diagnosis offered allows Putnam’s proof to fare better.

First, a refinement. The essence of the examples considered is that they involve advancing past a ‘tentative disjunction’. In the zoo and wall cases, for instance, the disjunctions would be something like

Either these animals are zebras or they are mules painstakingly and skilfully disguised as zebras

and

Either that wall is red or it is a white wall cleverly illuminated by concealed lighting to look as if it is red

Correspondingly, in the McKinsey case, the disjunction we have implicitly been working with would be:

Either I believe that water is wet or the seeming-thought which I attempt to express by “I believe that water is wet” is content-defective owing the reference failure of the purported natural kind term, “water”, in my language

But now an awkwardness surfaces in satisfying the conditions for the template: how is the disjunction relevantly tentative if in the formulation of its left-hand disjunct it avails itself of the very conceptual resource—the concept water—whose existence it was meant to be tentative about? Does not the disjunction go past the relevant issue no less than the simple claim that I believe that water is wet?

The simplest response to the difficulty is to introduce a concept, disjunction* and corresponding concepts of the other truth-functional connectives—which will allow the truth of a disjunct to suffice for the truth of a disjunction irrespective of the status—true, false or externally content-defective—of the other disjunct. With the occurrence of “or” so interpreted, the above formulation of the McKinsey disjunction may stand, the original argument now being construed as an enthymeme for the following


(0)\textsuperscript{M} Either I believe that water is wet *or* the seeming-thought which I attempt to express by "I believe that water is wet" is content-defective owing the reference failure of the purported natural kind term, "water", in my language

(1)\textsuperscript{M} I believe that water is wet

(2)\textsuperscript{M} Any thinker who believes that water is wet (belongs to a speech community which) has had such-and-such encounters with water \textsuperscript{20}

Therefore

(3)\textsuperscript{M} I (or my community) have had such-and-such encounters with water.

The transmission-questioning thought, to remind you, is then that, in advancing from (0)\textsuperscript{M} to (1)\textsuperscript{M}, the thinker presupposes a rational entitlement to discount the uncongenial right-hand disjunct, which in turn demands a prior rational entitlement—in the presence of (2)\textsuperscript{M}—to discount the negation of—i.e., affirm—(3)\textsuperscript{M}. So the argument from (0)\textsuperscript{M} to (3)\textsuperscript{M} is not at the service of a rational first conviction of (3)\textsuperscript{M}.

Why doesn't the Putnam argument fail of cogency in a similar way? Well: what is the appropriate tentative disjunction? It cannot be

(0)\textsuperscript{P} Either I have the concept of what it would be to be a brain-in-a-vat (in Putnam's scenario, etc.) or the seeming-thoughts which I attempt to express by tokenings of "brain-in-a-vat" are content-defective owing the reference failure of the term, "brain-in-a-vat", in my language

For there is no suggestion—I mean: no relevant suggestion—that envathood would generate illusions of content.\textsuperscript{21} The working assumption of Putnam's whole dialectic is that it is an epistemic possibility that the brains-in-a-vat could think: one is to start out open minded about whether these thoughts—one's own thoughts—might after all be the thoughts of a brain-in-a-vat (that they might be is exactly the Sceptic's teasing worry) and work to the conclusion that, such is their content, they couldn't be. What envyhooth, in Putnam's attendant scenario, would generate, rather—according to the relevant type of externalism—is the possession of a different conceptual repertoire to that possessed by normal humans in the world as we take it to be. So what is wanted is a disjunction which somehow expresses caution with respect to the
identity of the concepts one actually possesses, rather than their existence. It is exactly on this score that the natural intuitive reservation—the one almost everyone feels on first encounter—about Putnam’s proof occurs: “Couldn’t a brain-in-a-vat follow through a routine of thought-tokens isomorphic to Putnam’s proof—so how do I know that that is not what is happening when I do it?” (or something along those lines).

There is, however, no way to formulate such a disjunction. To see this, reflect that—when defective content is not an issue—we cannot make use, in either disjunct, of any concepts which would be exclusively available to a normal human or brain-in-a-vat respectively. For in that case, rather than being compatible with caution, a thinker’s endorsement of the disjunction would be a commitment to her having one status in particular or—incoherently—to her having both. It follows that the premise of Putnam’s proof in MC form—that I have the concept of what it would be to be a brain-in-a-vat (in Putnam’s scenario, etc.)—cannot be a disjunct of a suitable such disjunction. Remember that we are not at this point doubting the validity of Putnam’s proof—the issue concerns its cogency—and the proof does validly deliver that only a non-brain-in-a-vat can have that concept. If, on the other hand, we observe this restriction, using in the specification of the disjuncts only neutral concepts, as it were, then neither disjunct will entail that I am not a brain-in-a-vat (or at least, to be completely safe, neither can entail that proposition in Putnam’s way, just by exploitation of the conceptual resources it involves). It follow that nothing like the template for transmission failure operative in the McKinsey argument can engage here. That template demands a disjunction whose disjuncts would respectively be verified by distinct albeit subjectively indistinguishable states, although only one of the disjuncts suffices for the original conclusion. There is simply no candidate for a suitable tentative disjunction in the Putnam case.

This point does not force one to deny, of course, that the state of consciousness of a brain-in-a-vat might in certain relevant respects be subjectively indistinguishable from that of a normal human possessed of the concept, brain-in-a-vat. Since the state of consciousness of a thinker who had never encountered water might in relevant respects be subjectively indistinguishable from that of a normal earthly subject possessed of the concept, water, there can still be that much analogy between the cases. But the crucial point for the claim of transmission failure is the charge that in helping oneself to the premise(s) of the targeted argument one
has passed over a different, thinkable possibility which would obtain in circumstances subjectively indistinguishable from those one finds oneself in. In the McKinsey case, that is the possibility that one’s token thoughts involving “water” are empty of real content. If the relevant strong externalism is true, and certain types of sceptical thought about one’s past are also genuine possibilities, then that will indeed be a thinkable possibility. But in the Putnam case, what has to be subjectively opaque to one is not the existence but the identity of a certain concept: there has to be a thinkable issue about which concept one tokens by “brain-in-a-vat”. The key to understanding Putnam’s proof is to see that it fineses any such issue. Suppose a philosopher thinks there is such an issue—she still has to grant that, though the expression might of course have denoted a different concept, there is no thinkable alternative to the claim that by uses of “brain-in-a-vat” one actually tokens the concept: brain-in-a-vat, (add sotto voce, if you like: “whichever concept that is”). That’s just a use of disquotation. And the question which Putnam’s argument addresses—and validly settles in the negative—is whether I myself fall under that concept (sotto voce: “whichever concept that is”). Any issue of the transcendental identity of the concept—supposing there is indeed any intelligible such issue—simply doesn’t impinge on the proof at all.

So we have found a disanalogy. If Putnam’s proof, though valid, fails of cogency it seems it must do so for reasons quite different to those operative in the McKinsey case. My view remains that it is cogent. But I do not, in saying that, withdraw the various claims about its limited significance, both metaphysical and epistemological, which I have made in other work.22, 23

Notes


3. This regimentation is borrowed from Martin Davies' "Externalism, Architecturalism and Epistemic Warrant", in C. Wright, B. Smith and C. MacDonald, eds., op. cit. n. 2, pp. 321–61. So far as I am aware, Davies was the first to raise in print an issue about the transmission of warrant in the McKinsey argument. I shall return to his discussion below.

4. It does not matter whether either of the hereby mooted forms of this premise are especially plausible. The crucial point is that a paradox of the relevant kind will be generated by any form of externalism which sets definite a priori historical preconditions for concept possession.

5. But not unexampled. The interested reader should consult Boghossian’s op. cit. n. 2 section VI.

6. Martin Davies, op. cit. n. 3 above. Davies draws on the discussion of transmission failure which I applied to Moore’s so-styled ‘proof’ of an external world in my “Facts and Certainty”, (Henriette Hertz Philosophical Lecture for the British Academy, December 1985, in Proceedings of the British Academy LXXI, pp. 429–72). But as he rightly stresses, the diagnosis of how transmission failures may arise which I offered there will not directly apply to the McKinsey argument. The treatments of that argument which we respectively offer, I below, and he both in the essay cited and—with significant improvements and changes—in his more recent “Externalism and Armchair Knowledge” (in New Essays on the A Priori, Paul Boghossian and Christopher Peacocke, eds., Oxford: The Clarendon Press, 2000) are very different in detail. The latter paper came to my attention only after I had completed this one. I hope to return to a detailed comparison and evaluation of Davies’ most recent proposals and my own in subsequent work.

7. I paraphrase. See Davies, op. cit. n. 3 above, p. 353.

8. Ibid., p. 354

9. However, see Mark Sainsbury’s observations in his Comments on this paper below; this volume, pp. 191–200.

10. I prescind, in this remark, from consideration of Davies’ recently refined version of (D) in his “Externalism and Armchair Knowledge”.


13. The details matter. Putnam’s scenario is one where the universe has always consisted just of brains-in-a-vat attended by unfailing automatic machinery, beyond which the void.

14. For a review of some such objections, and responses to them, see my op. cit. n. 12 above.


16. I borrow these well-known examples, of course, from Dretske op. cit. n. 1 above. However Dretske presents these cases as failures of closure.

17. This is of course a closure step.

18. Though it may well be adaptable to the information-dependence cases canvassed earlier; see Alfonso Suarez’s Comments below (this volume, pp. 164–71).
19. Another would be to essay metalinguistic formulation of the first disjunct.

20. Arguably, of course, the putative a priori warrant for \( (2)^M \) likewise goes past an uncongenial disjunct*—specifically:

the seeming-thought which I attempt to express by “Any thinker who believes that water is wet (belongs to a speech community which) has had such-and-such encounters with water” is content-defective owing the reference failure of the purported natural kind term, “water”, in my language.

But I spare the reader the complication of incorporating that into the reconstructed argument. Its presence would make no important difference; the effect would be merely that a prior rational entitlement to \( (3)^M \) would be likewise a presupposition of discounting that disjunct* and advancing to \( (2)^M \).

21. I don’t mean that there is no question but that Putnam’s brains-in-a-vat could possess concepts. The point is merely that this is not in question in Putnam’s dialectic. But if it is made an issue, then I would contend that the Putnam and McKinsey arguments are in the same case.

22. See my op. cit. n. 12 above. These claims were respectively that the argument was powerless both against sceptical arguments whose distinctive scenarios are careful to leave in place the external preconditions of the concepts involved—for instance, the scenario of envatment in my sleep last night, or the scenario that I am dreaming right now (which is all that is needed for the kind of version of the sceptical argument from dreaming canvassed e.g. by Barry Stroud—see e.g. ch. 1 of his The Significance of Philosophical Scepticism, Oxford: The Clarendon Press, 1984—and myself—see “Scepticism and Dreaming: Imploding the Demon”, Mind C (1991), pp. 87–116); and that the bearing of Putnam’s proof is in any case restricted to sceptical scenarios whose detail is available to us; it is powerless to disrupt the general thought—integral to metaphysical realism if anything is—that we might be so situated in the world that while insured, for externalist reasons, against possibilities of massive error, we remain massively ignorant, because barred from the concepts we need to capture fundamental aspects of the world and our situation in it. That is exactly the predicament of the brains-in-a-vat in his scenario; and the most the proof yields against it is that if that were our predicament, we would not be able to understand a description of its detail. The coldly alien, potentially impenetrable world of metaphysical realism thus looms no smaller, so to speak, after the argument than before. Of course, Putnam has resisted this suggestion (see his “Comments and Replies” in Clark and Hale, eds., op. cit. n. 12 above). But I must defer a response to him to another occasion.

23. This paper is a development of one section of more general work in progress of mine on the acquisition of warrant by inference. It has benefited from discussion of the latter at seminars at Aarhus University, Berkeley, Edinburgh, New York University, Oxford, Princeton and Yale, and from a self-contained presentation at the one-day conference on Externalism and Self-knowledge held at the University of Bristol in February 1999. My thanks to the participants on all those occasions and especially to John Campbell and Tim Williamson for detailed responses, and to Lars Gundersen, Duncan Pritchard and Sven Rosenkranz for much helpful discussion.