ROBUST BELIEF STATES
AND THE RIGHT/WRONG DICHOTOMY

Robert Stainton

1. INTRODUCTION

A person who believes that *p*, when *p* is false, goes wrong somehow. Whereas that same person, who believes that *p* when *p* is true, "gets it right". Some philosophers, purportedly inspired by Wittgenstein (1953), have supposed that this "central fact" about believing shows that beliefs are not states — at least not in one familiar sense of the word 'state'. In what follows, I consider one variant of this style of argument. It goes like this:

1. *The Quasi-Wittgensteinian Argument*

Premise 1: Necessarily, being in any belief state makes the believer either right or wrong.
Premise 2: If beliefs are robust states, then it's not the case that, necessarily, being in any belief state makes the believer either right or wrong.
Conclusion: Beliefs are not robust states.

After explaining the argument, I suggest ways in which it may be resisted.

Precisely what it means to say that beliefs are not "robust states" is not immediately clear. Not least because the characterization varies from one author to another.[1] But the general idea is this. There is an intuitive contrast between states like having mass *m*, or electrical charge *c*, and states like having certain rights, or being the winner of a game. The former are intrinsic, objectively observable, causally efficacious, spatially and temporally located; they admit of scientific theorizing, and can be captured by descriptive physical laws; and so on. The latter are none of these. Putting aside the question of what *exactly* the difference is, call the former "robust" or "inflated" states. The latter, if one is to call them states at all, may at best be termed "deflated" states.

---

[1] Two authors who pursue this sort of line are Arthur Collins (1979, 1987, 1994, 1996) and Norman Malcolm (1991). However, the exact argument that I consider below cannot be fairly attributed to either of them.

The state of believing belongs in the second category, if the above argument is correct. As one says: to believe something is simply to occupy a certain role in a game — the central feature of this game being that believers *lose* when what they believe is false, and they *win* when what they believe is true. To believe that *p* is not, therefore, to be in a spatially isolable, causally efficacious, or physically constituted state. Or again, as Arthur Collins (1996, 319) puts the point: beliefs are not "internal explanatory realities".

To repeat: I will be not be concerned with clarifying this (rather obscure) conclusion. Suffice it to say that, if it is correct, many standard views in philosophy of mind are questionable: both the Identity theory and Functionalism, for instance, are committed to beliefs either being, or being constituted by, physical states with mass, location, causal powers, etc. Instead of elucidating the conclusion, I want to discuss the aforementioned argument for it — suggested by, though (so far as I know) not proposed in, the Wittgenstein literature.

2. AN INITIAL REPLY TO THE "QUASI-WITTGENSTEINIAN ARGUMENT"

Let me start by granting the argument may seem initially plausible. There are only two premises, and each has significant appeal. Premise 1 follows from the "central fact": epistemic success or failure is the very nub of believing, not some merely contingent feature of it. Premise 2 is initially supported by contrasting the analysis of $\ulcorner S$ believes that $p \urcorner$ with $\ulcorner S$ is in $N \urcorner$. While the former makes essential reference to *p*, the sentence $\ulcorner S$ is in $N \urcorner$ has a subject matter wholly distinct from *p*, for any inflated state *N* — as we can see, the story goes, by the fact that it is always at least intelligible to say $\ulcorner N$ is present in me, but I do not take any stand on $p \urcorner$, whereas it's quite absurd to say $\ulcorner$I believe that *p*, but I do not take any stand on $p \urcorner$. It is the intelligibility of the former which (purportedly) shows that if beliefs are robust states, then it's not the case that, necessarily, being in any belief state makes the believer either right or wrong. For, given its intelligibility, the following is also conceivable: $\ulcorner$Robust state *N* is present in me, and *p* is false, but it is not the case that I am mistaken about $p \urcorner$. Which seems to show that *N*'s being present brings with it no right/wrong dichotomy with respect to *p*. Hence, it may seem, for any "robust" state *N*, being in *N* cannot be sufficient grounds for a person being mistaken: whether *N* is a neurological, functional, or other state, one could be in *N* and not be either correct or mistaken about anything. Because, to repeat, it's perfectly possible to suppose that one is in *N*, while simultaneously withholding any view on *p*. (See Collins (1994, 931) for discussion.)

One might complain that this defense of P2 confuses conceptual necessity with necessity *tout court*. Maybe the situation is this: though one can intelligibly imagine oneself in *N*, while not being at "epistemic risk" vis à vis *p*, being in *N* while not being beholden to *p*'s truth isn't really possible. (Compare: maybe one can intelligibly conceive of mass without energy, but that

doesn't mean there can be mass without energy.) While I have real sympathy with this complaint, that is not the line I will pursue right now. (Something similarly will, however, emerge at the end.) For now, I want to consider a reply which focuses on the validity of the argument. So, here goes.

Grant that both premises are true, when read in the right way. But, goes this initial reply, there is an equivocation, between the first and second premise, on the scope of 'any belief state'. Specifically, to make both premises plausible, 'any belief state' must be given *narrow* scope in Premise 1, and *wide* scope in Premise 2, relative to the modal adverb. Thus, (2) and (3) are plausible — but, reading the Premises as (2) and (3) respectively, the argument is invalid: these two don't entail that beliefs aren't robust states.

2. $\Box \forall x$ [Belief(x) $\supset$ $\forall y$ (Agent(y) $\supset$ (In(y)(x) $\supset$ (Right(y) $\vee$ Wrong(y))))]
3. If beliefs are robust states, then $\sim\forall x$ [Belief(x) $\supset$ $\Box\forall y$ (Agent(y) $\supset$ (In(y)(x) $\supset$ (Right(y) $\vee$ Wrong(y))))]

On the other hand, (2) and (4), taken together, really *would* defeat inflationary views of belief — because 'any belief' is now uniformly given narrow scope.

4. If beliefs are robust states, then $\sim\Box\forall x$ [Belief(x) $\supset$ $\forall y$ (Agent(y) $\supset$ (In(y)(x) $\supset$ (Right(y) $\vee$ Wrong(y))))]

So, if the Premises are read as (2) and (4), the argument becomes valid. The problem is, (4) isn't *prima facie* plausible: surely, one wants to say, beliefs *so described* make believers either right or wrong, *even if* beliefs are robust states.

Here's a rough and ready illustration of the scope difference. Inflationism about beliefs might well be committed to (5), because the very (neural, functional, etc.) state which *actually* is the belief state might not have been a belief state at all. But inflationism surely needn't be committed to (6) — in which the modal 'there's a possible world' has wide scope over the description.

5. The entity which *in fact* is the belief state that it snows in Montevideo is such that there's a possible world where it does not make its possessor right or wrong.
6. There's a possible world where the belief state that it snows in Montevideo, whatever it's neural/functional/etc. realization, does not make its possessor right or wrong.

So, either the argument in (1) is fallacious — because of a scope equivocation — or one of its premises lacks *prima facie* plausibility. Either way, the argument fails.

3. REPLY TO THE SCOPE-BASED REPLY

This scope-based response may face an interesting obstacle, if one takes ⌜is the belief state that p⌝ to be a natural kind term. Assuming Kripke (1971, 1972) is right, a natural kind term denotes the same kind in every possible world. (See Putnam 1973, 1975 for related remarks.) But then, if ⌜is the belief state that p⌝ is a natural kind term — as those who posit robust states likely assume, in so far as they believe that there could be a *science* of psychology — it would seem that every actual belief is *necessarily* a belief. How does this affect the scope-based reply to (1)? Well, by adding (7) to the wide scope reading of Premise 2 (i.e. (3)), its narrow scope reading (namely, (4)) also becomes ever so plausible.[2]

   7. $\forall x$ [Belief(x) $\supset$ $\Box$ Belief(x)]

That is, the anti-inflationist can say something like: "I read Premises One and Two as unequivocally assigning *narrow* scope to 'any belief'. Hence the argument is valid. Importantly, however, Premise Two, even read in this way, is nevertheless plausible — because it follows from (3) and (7), each of which is granted to be independently tenable."

   It turns out, though, that an anti-inflationist cannot make this move: he cannot avoid the equivocation-or-implausibility problem in this way. The thing about natural kind terms is, one discovers their satisfaction conditions only through empirical (especially scientific) investigation. So, *if* ⌜is the belief state that p⌝ rigidly designates a natural kind, then one cannot discover the truth conditions of ⌜believes that p⌝ from the armchair — because armchair semantic investigation is wholly *a priori*. (And recall that what initially supported P2 was an *analysis* of 'believes that'.) At best, one can "discover" how the reference of 'belief state' gets fixed. Indeed, maybe it's fixed as follows: Let ⌜is the belief state that p⌝ rigidly designate that actual state, the presence of which makes S mistaken if $\sim p$. In which case, there are no longer any grounds for urging that premises one and two are true: How could anyone distinguish the essential from the contingent properties of belief states, on the basis of semantic intuitions, if belief states are natural kinds — whether assuming inflationism or not?[3]

---

[2] As Crispin Wright kindly pointed out to me, what the defender of the argument really needs is that ⌜is the belief state that p⌝ be a rigid designator. One way to get this result is if it denotes a natural kind. And this is the route I explore, because there's a certain initial plausibility to the idea that, for the "robust" theorist, there can be a science of belief states — which, in turn, yields that they are natural kinds. In the end, as I'll shortly explain, this line of thought does not pan out. Of course, if there is another argument to the conclusion that ⌜is the belief state that p⌝ is a rigid designator, which does not appeal to natural kinds, then this "reply to the reply" will need to be revisited.

[3] In distinguishing between giving the meaning and describing how reference is fixed, I of course follow Kripke, who writes: "...the name [Cicero] would not be synonymous

Besides, however unclear the anti-inflationist conclusion, it surely is *not* consistent with the idea that beliefs are natural kinds! For natural kinds are "internal explanatory realities" *par excellence.* So, the anti-inflationist, in defending (1), really ought not dismiss the equivocation (i.e. between a wide and narrow scope reading for 'any belief state') by appealing to the subtleties of natural-kind-terms semantics. Which leaves the proponent of (1) stuck with the problem previously stated: Premise 1 is plausible only when 'any belief state' is given narrow scope, whereas Premise 2 is plausible only when 'any belief' is given wide scope.

4. THE EPISTEMIC RELEVANCE OF ROBUST STATES

Let me consider one final move which the anti-inflationist might make, to salvage the argument in (1). Premise Two isn't plausible, when 'any belief' is given *narrow* scope with respect to the modal 'necessarily': I repeat, it's surely the case that, even granting inflationism about beliefs, beliefs-so-described make one right or wrong. Premise One, on the other hand, isn't plausible when given *wide* scope. Or is it? It might be suggested that there's *something* right about Premise 1, even when the quantifier phrase 'any belief state' is given wide scope over the modal operator — though it's doubtful whether □ is the right modality. In place of □, one ought only demand, for example, that being in any belief state, however described, be *relevant* to whether the agent is either right or wrong. One state (or state-of-affairs) necessitating another is a quite strong modal relationship; but one state (or state-of-affairs) *being relevant* to another is a comparatively weak modal relationship. Weak enough that something like Premise One, with wide scope 'any belief', may yet be rendered plausible — e.g. being in any belief state, however described, is *relevant to* the believer being either right or wrong. (I fear that the term 'relevant to' is not truly felicitous, but I'm unable to think of a better word. What I have in mind is the weakest modal relationship such that, plausibly, it does obtain between ⌜believes that *p*⌝ and ⌜is right/wrong⌝ and it doesn't obtain between ⌜is in robust state *N*⌝ and ⌜is right/wrong⌝. If there is

_____

with the description, but would be used to *name* an object which we pick out by the contingent fact that it satisfies a certain description" (1971, 185). As Kripke notes, the same consideration applies, *mutatis mutandis*, to 'heat' and other natural kind terms: 'heat' doesn't *mean* sensation thus-and-such. Rather, 'heat' rigidly designates molecular motion of a certain kind – though this reference gets establish via the description 'sensation thus-and-such'.

Of course the anti-inflationist might say that he *does* have grounds for Premise One, on its narrow scope reading: from how the reference of 'belief state' is fixed. (I.e., it is the *a priori* knowledge of how reference is fixed which rules out inflationism.) But, taking seriously the idea that the anti-inflationist is (at best) describing reference fixing for 'belief state', and not giving its "meaning", the narrow scope reading of Premise One is precisely as plausible as its wide scope reading. Which is to say, it's not plausible at all — at least where the modality is □.

such a modal relationship, then the argument introduced at the outset can be restated, but substituting the name of this relation for 'necessarily'.)

Put symbolically, introducing a modal connective **Rel** for '__ is relevant to __', Premise One should be replaced by:

8. $\forall x \,\forall y$ [(Belief(x) & Agent(y)) $\supset$ [In(y)(x) **Rel** Right(y) $\vee$ Wrong(y)]]

For Premise Two, one may then substitute (9):

9. If beliefs are robust states, then $\sim\!\forall x \,\forall y$ [(Belief(x) & Agent(y)) $\supset$ [In(y)(x) **Rel** Right(y) $\vee$ Wrong(y)]}

Crucially, this recasting of Premise Two also gives 'any belief' wide scope over the modal connective — so *both* Premises are interpreted with 'any belief' having wide scope.[4] Hence there's no equivocation. And, as in the $\square$ version, revised Premise Two, read with wide-scope 'any belief' plus a weaker modality, remains plausible: that is, it's not obvious that, according to inflationism, being in a belief-state-however-described is even *relevant* to being right or wrong.[5] So there may yet be a problem for robust belief states.

An example may make the move clearer. It seems that if inflated state *N* really *is*, for example, the belief that Cobain is alive, then its presence ought to be *modally connected*, somehow or other, to the proposition that the believer is subject to being correct/mistaken. But, asks the anti-inflationist, how can the presence of a functional, or neural, state in me *have such a modal connection* to, for example, my being right if it's raining, or my being wrong if Cobain is dead? Functional states, neural firings, etc. seem to be matters of fact wholly unrelated to such normative claims. So, returning to my example, if inflationism is true, then Cobain's being dead won't even be *"relevant"* to evaluating the believer's rightness/wrongness.[6] Which suggests that beliefs cannot be robust states.

_____

[4] Strictly speaking, the notion of "scope" is different here. After all, **Rel** is a *modal sentential connective*: it forms modal sentences from two (open or closed) sentences. Thus **Rel**, unlike $\square$, is not a sentential operator. In the case of **Rel**, then, the notion of scope at play is comparable to a quantifier having scope *only* over the antecedent of a conditional, rather than over the whole sentence. Whereas scope, for $\square$, is roughly parallel to scope of negation.

[5] Moreover, *this* wide scope rendering of Premise Two is plausible, *even if beliefs aren't natural kinds*. Or so I'm willing to suppose for the sake of argument.

[6] Or again, consider another argument extrapolated from Collins 1987. First of all, (i) is true only if (ii) is.

i) ⌜*S* believes that *p*⌝
ii) ⌜*S* is right about *p*, or *S* is wrong about *p*⌝

Now, each disjunct of (ii) requires a truth value for *p*. In which case, it is a necessary condition for the truth of (i) that *p* have a truth value. So far so good. Now, is there an

What this shows, however, is not that beliefs cannot be inflated states; instead, it confirms that beliefs, if they are inflated states, must be states whose existence *is* "relevant" — i.e. modally connected to — the truth or falsity of certain normative claims. That is, inflationists cannot accept (9). Indeed, the falsehood of (9) must be insisted upon as a constraint on the nature of (inflated) beliefs. In which case, to take an example, any inflated state which makes 'John believes that Cobain is alive' true must be such that being in that state *is* somehow non-accidentally related to Cobain's being alive or dead. On the other hand, to return to a point noted in the previous section, one need not presume that this modal connection will be introspectively available, or a matter of semantic analysis. That the modal connection exist non-accidentally is required; that this requirement be met by a conceptual link is quite another matter.

But, crucially, the representational theory of mind meets this twin demand. For, in RTM, the sentences which are tokened in the believer's brain have both syntax *and semantics*. Thus, to take an example, if B_{2765} being in John's "belief box" constitutes his belief that Cobain is alive, then its presence, together with the fact that Cobain is dead, *will* be relevant to judging that John is mistaken — because, *ex hypothesis*, B_{2765} means that Cobain is alive.[7] Crucially, however, that B_{2765} *has* meaning, and what it's meaning is, is introspectively available at best under some descriptions: essentially because meaning-facts about B_{2765} are not conceptual truths — rather, they depend on who-knows-what psycho-physical laws. Or so the RTM story goes. So, the right way around this final argument is to bite the bullet: accept (8), but deny that inflationism is inconsistent with it. That is, epistemic success or failure really is essential to belief — but this poses no problem for robust belief states, when properly construed.

Time to sum up. I began by introducing (1), a "Quasi-Wittgensteinian Argument" against treating beliefs as robust states. I criticized the argument, on the grounds that *either* it suffers a scoped-based equivocation, *or* its second premise is implausible. (If the former, the argument is invalid; if the latter, it (likely) has a false premise; either way, it's unsound.) I then offered two replies, on behalf of (1). The first made the argument unequivocal by having 'any belief state' take narrow scope in both premises. It then appealed to the semantics of natural kind terms, in order to defend the plausibility of Premise 2 — even on its narrow scope reading. The second avoided equivo-

---

inflated state *N* of *S,* such that *S*'s being in *N* has *any bearing* on whether *p* has a truth value? No, says Collins: *S*'s being in *N*, no matter what *N*'s constitution, is immaterial to whether *p* has a truth value.

[7] Arthur Collins misses this crucial feature of the representational theory of mind because he steadfastly refuses to admit that datable concreta, including mentalese tokens, can have truth conditions. He supposes, for reasons which remain obscure, that only abstract propositions can be true or false. (See Collins 1979 for discussion.) That this is an error is shown by the very fact that the word tokens on this page are truth evaluable.

cation by giving 'any belief state' wide scope in both premises, while weakening the modality, to make Premise 1 plausible. Neither defense of (1) was effective, however. The narrow-scope defense foundered on two features of natural kinds: treating beliefs as natural kinds renders them too "inflated", and conflicts with the premise-supporting appeal to *a priori* intuitions about truth conditions, essences and such. The wide-scope defense, on the other hand, because it must weaken the modality in *both* premises one and two, makes Premise 2 quite implausible — at least if one assumes a representational theory of belief. (It's part and parcel of representational theories that belief states are contentful, and that the content they have is no mere accident.) The overall conclusion then, is this: once "robust beliefs" are properly construed, it is no skin off the inflationist's nose that a person who believes that *p*, when *p* is false, has failed epistemically.[8]

Robert Stainton
Dep. of Philosophy
1125 Colonel By Drive
Ottawa, ON K1S 5B6 Canada
stainton@ccs.carleton.ca

References

Collins, Arthur W. (1996) "Moore's Paradox and Epistemic Risk" in *The Philosophical Quarterly,* 46, 308-319.
—— (1994) "Reply to Commentators" in *Philosophy and Phenomenological Research,* 54(4), 929-945.
—— (1987) *The Nature of Mental Things.* Notre Dame: University of Notre Dame Press.
—— (1979) "Could our beliefs be representations in our brains?" in *The Journal of Philosophy,* 76(5), 225-243.
Kripke, Saul (1972). *Naming and Necessity.* Cambridge: Harvard University Press.
—— (1971) "Identity and Necessity" in M.K. Munitz (ed.) *Identity and Individuation.* New York: New York University Press. Reprinted in A.W. Moore (ed.)(1993) *Meaning and Reference.* Oxford: Oxford University Press.
Malcolm, Norman (1991) "I Believe that *p*" in E. LePore and R. van Gulick (eds.) *John Searle and his critics.* Oxford: Blackwell.
Putnam, Hilary (1975) "The meaning of 'meaning'". In K. Gunderson (ed.) *Language, Mind and Knowledge.* Minnesota Studies in the Philosophy of

---

Science, vol. VII. Minneapolis: University of Minnesota Press. Reprinted in H. Putnam (1975) *Mind, Language and Reality.* Cambridge: Cambridge University Press.

—— (1973) "Meaning and Reference" in *The Journal of Philosophy,* 70, 699-711. Reprinted in A.P. Martinich (ed.)(1990) *The Philosophy of Language.* Oxford: Oxford University Press.

Wittgenstein, Ludwig (1953) *Philosophical Investigations.* Oxford: Blackwell.