ABSTRACT. A chronic difficulty for functionalism is the problem of instantiations of a functionalist theory of mind which seem to lack some or all of the mental states—especially qualitative—we want to attribute to minds the theory describes. Here I discuss one such counterexample, Block's system S, consisting of the population of China organized to simulate a single mind as described by some true, adequate, psychofunctionalist theory. I then defend a version of functionalism against this example, in part by an adaptation of Dennett's notion of "stances". A true, adequate theory, as Block understands it, would be appropriate to Dennett's "design" or (at best) "intentional" stance; but a genuinely true and adequate theory should instead coincide with a "personal" stance. Hence, if system S does instantiate such a theory, we must impute to it mental states, even qualitative, whether or not it "really" has them. Hence Block's counterexample lacks force.

Ned Block's article, "Troubles with Functionalism"[1], offers several examples against several varieties of functionalism. Here I shall discuss one of those examples, that of the population of China simulating a single mind, as it applies to what Block calls "psychofunctionalism."

For Block, psychofunctionalism identifies a (type of) mental state, A, with A's Ramsey functional correlate with respect to some scientific (as opposed to a common sense) psychological theory.[2] The Ramsey functional correlate is spelled out in terms of that theory, which characterizes causal structures of inputs and outputs related to variables which range over mental states (perhaps among other
things). [3] Two systems, \( X \) and \( Y \), are to be regarded as functionally equivalent, if and only if they are equivalent with respect to at least one (reasonably adequate, true) psychological theory[4]—in other words, just in case they both can be characterized by the same set of causal structures of inputs, outputs, and variables.

Now, according to Block, a billion people, e.g., the population of China, could conspire to form a system, \( \mathcal{S} \), that was functionally equivalent to a single human mind, \( M \). But the contrived system \( \mathcal{S} \) could not have certain mental states which mind \( M \) has, and hence the "adequate, true" psychofunctional theory involved is inadequate to account for the presence of those mental states in systems which do have them.

First, I shall characterize the example as it applies to psychofunctionalism. Then I shall construct a defense for psychofunctionalism against this example, by means of some of Dennett's ideas.

1. THE PROPOSED COUNTEREXAMPLE

We are asked to suppose that: the population of China are organized into an interactive system, \( \mathcal{S} \), such that some (true, adequate) psychological theory (relating inputs, outputs, and variables by means of Ramsey sentences) of a single mind, \( M \), is also true of \( \mathcal{S} \). Apparently, one human or one committee of humans in \( \mathcal{S} \) corresponds to one mental structure in \( M \). There is some way of characterizing inputs and outputs such that both \( M \) and \( \mathcal{S} \) and their respective internal components (mental structures in \( M \), people in \( \mathcal{S} \)) can be described as receiving the same inputs and producing the same outputs. By hypothesis, \( M \) and \( \mathcal{S} \) are (psycho) functionally equivalent.

Apparently, \( \mathcal{S} \) could have some of the (nonqualitative) mental states ordinarily ascribed to minds such as \( M \). If \( M \) could recall the words to a song, believe it to be Friday morning, solve an equation, or desire to know the way to San Jose, so, by hypothesis, could \( \mathcal{S} \). Since (by hypothesis) \( \mathcal{S} \) performs these functions in the same way—using the same steps involving the same internal units to process the same information—as \( M \), it is reasonable to say that \( \mathcal{S} \) goes through the same mental states as \( M \). It is immaterial that \( M \) does it through configurations of neural activity in certain sequences and coordinations while \( \mathcal{S} \) does it through (corresponding) configurations of human activity (e.g., "storing" and "retrieving" notes from files, passing notes to one another); after all, computers are credited with going through some mental states even though they do it as configurations of electrical activity.[5]

Nevertheless, Block thinks it doubtful that \( \mathcal{S} \) has
One of his reasons for this is that he does not believe psychological theories are designed to explain qualia. Where mental life has a qualitative aspect and is explained, e.g., the case of a type of sensation, it is the functional aspect and not the qualitative that is explained.\[6\]

But suppose we grant Block this point: that psychological theories do not explain qualia. Might it be true nonetheless that any system (psycho)functionally equivalent to a mind does have qualitative mental states?

2. AN UNSUCCESSFUL REPLY

Anyone who can imagine Block's system $S$ of a billion people organized to be functionally equivalent to a single mind $M$, is likely nonetheless to be inclined to suppose that, while he could interact intellectually with this system in the same ways he could with another human or with a (well-programmed) computer, he would not be inclined to impute to it such qualia as, e.g., pains. But might this be because of our attitude towards the components, rather than to the nature of system $S$? Block considers a reply in the spirit of Putnam,\[7\] offering the requirement that no system capable of such qualia as feeling pain "possess a decomposition into parts which separately possess" these qualia. Since any one of the Chinese of which $S$ is composed is eminently subject to such qualia, system $S$ is disqualified on this ground, without thereby undermining psychofunctionalism (as so amended). The intuition this reaches for is that the component cannot have the quality for the whole system. Surely, as Block points out, if what we now regard as elementary particles were discovered to consist of sentient beings, this would not undermine our belief that we, who are composed of them, experience pain, etc. These tiny beings would play the same role in a lump of carbon that they would in a human. They do not experience our pains for us.\[8\] Yet surely, were we to dismantle a presumptive human being and find inside a midget pulling levers, monitoring screens, and so on, we would on this basis deny that the original had any qualia.

Does this requirement, inspired by Putnam, eliminate Block's system $S$ as a counterexample? It does not, since any of the humans composing $S$ corresponds to some theoretical component of $M$. If $M$ has a pain-experiencer component then the proposed requirement is simply false: $M$ has qualia, by hypothesis, so if its component had them too, $M$ itself would be eliminated. If $M$ has not such components then the humans in $S$, as corresponding to these components, do not have what qualia they have for the system $S$, but have them only for themselves.
3. A MORE PLAUSIBLE REPLY

Another variant of Putnam's suggestion is more plausible: A system may have qualia, but we cannot simultaneously regard both it and any component of it, as system and component, as having the same qualia.

Hence were elementary particles discovered to consist of sentient beings, we would think of them, as components of some human being (or lump of carbon) as elementary particles, even though we could separately think of them as sentient beings having the same qualia as the human body in question. Of course, both the human and the tiny beings would actually have these qualia, despite our limited ability to regard them so. The dummy manipulated from inside by midget does not have qualia in the first place, but certainly any inclination to think him to vanishes upon our discovery of the midget within.

This brings us to the problem of how to decide whether Block's system $S$ does indeed possess qualia in addition to nonqualitative mental states. If my proposed caveat is legitimate, we are not allowed to think of $S$ as composed of a billion human beings, in reaching our decision. Indeed, the obvious and most natural approach would be to consider $S$ as a whole, just as we would any apparently mental entity. It would seem that the appropriate question to ask is, what stance, in Dennett's sense, would be the best to take toward $S$ in interacting with it.

For Dennett there is an hierarchy of possible stances to take toward any system--whether computer, mere animal, human, or the contrived system $S$ at issue--in interacting with it. These stances are, in ascending order, the physical, design, intentional, and personal stances.[9] In principle, any stance can be taken toward any system, but in fact, given various contingent factors, there is with respect to a given system in a given situation one correct or best stance. The decision as to stance is to be made on the basis of success rate in prediction and explanation of the system's behavior according to the assumptions built into each of the possible stances.[10] It goes without saying that an adequate (and true) psychological theory of any system must incorporate the assumptions of the correct stance.

Block's claims are not framed in terms of stances. Nevertheless it is apparent that he would believe the intentional stance correct regarding system $S$. For this stance one assumes the system's rationality and predicts and explains in terms of intentionality:

One predicts behavior in such a case by ascribing to the system the possession of certain information and supposing it to be directed by certain
goals . . .

In taking this stance, one explains and predicts the system's behavior by ascribing beliefs and desires to it.\[11\]

As Block and Dennett both realize, this is something less than attributing to the system in question such qualia as pains. Hence were psychological theories appropriate to the intentional stance the truest and most adequate concerning all mental or quasi-mental systems, Block would be correct in his claim that an adequate, true psychological theory of mind could be instantiated by a system innocent of qualitative states. But the personal stance seems to me more appropriate than the intentional.

Dennett distinguishes two concepts of personhood, the metaphysical and the moral. The metaphysical is of "an intelligent, conscious, feeling agent;" the moral concept presupposes this and includes (moral, accountability and possession of rights and responsibilities as well. But in the end there is only one personal stance; the metaphysical and moral concepts are not really separate and distinct, but are "just two different and unstable resting points on the same continuum".\[12\]

Presumably M, as haver of such qualia as pain, must fit at least the metaphysical concept of personhood, thereby justifying a personal stance toward M. Hence for a psychofunctional theory to be true and adequate concerning M, it should be consonant with a personal stance, and S, to be psychofunctionally equivalent to M, must instantiate such a theory.

It might be objected on behalf of Block that Dennett lists six necessary conditions for personhood, i.e., for justifying a personal stance, and that the having of qualia is not listed among these. But in reply I note that such things as consciousness, including self-consciousness, and reciprocity of behavior are required.\[13\] Now it seems to me (though Dennett ignores this possibility) that qualia are intimately tied to these attributes (reciprocity, consciousness, self-consciousness). If M has them (pains etc.) then presumably they play a causal role in his behavior. This is true even if, as I have noted Block maintains, some mental states count both functionally and qualitatively and it is their functional aspect rather than their qualitative aspect which plays the causal role. M is still to be explained in part in terms of mental states which count as qualitative playing a functional role, where presumably they would not play the same role were they not these same states (which are qualitative). Now since by hypothesis S, the organization of a billion humans, is to be understood by means of the same theory as M, S too must be understood from the same personal stance, and as having the same qualitative mental
states as M. In other words: a true, adequate psychological theory of M might not explain M's qualitative states as such; nevertheless it uses those mental states which are qualitative (as well as play a functional role) to predict and explain M's behavior. Since S is psychofunctionally equivalent to M, S instantiates the same theory, the same predictions and explanations are made concerning S, and hence the same qualitative states must be imputed to S because of their explanatory and predictive role.\[14\]

Block believes to the contrary, that it is possible "that the . . . Psychofunctional state . . . Psychofunctionalists would want to identify with pain can occur without any quale occurring." His examples of this possibility are all along the lines of system S, which is of course in dispute, and hence cannot count as reliable evidence. He does offer an example in which it would seem qualia occur in the absence of the associated functional state:

After frontal lobotomies, patients typically report that they still have pains, though the pains no longer bother them. . . . These patients show all the sensory signs of pain (e.g., recognizing pin pricks as sharp), but they often have little or no desire to avoid 'painful' stimuli.

It is on the basis of such examples that Block suggests that each pain is either "a composite state whose components are a quale and a . . . psychofunctional state" or "a quale playing a certain . . . psychofunctional role".\[15\]

Despite Block's intent, the example of lobotomy patients counts rather in favor of my contention that mental states counting somehow as qualitative would not play the same (functional) role were they not the same qualitatively. Pain is somehow different for the lobotomized. It still counts as pain for them, but no longer the urgent, motivating sort. A Ramsey sentence with a variable instantiable, say, by a stabbing pain to the gluteus maximus, would associate this variable in a causal relationship with certain inputs and certain outputs. Perhaps the inputs would be the same for the normal and the lobotomized. But surely the outputs associated with that variable and those inputs according to the Ramsey sentence instantiated by the normal subject, are different from those for the lobotomized subject. The mental state of the normal subject is involved with output directed toward modifying or alleviating the painful situation (he is very likely to stand up quickly or jerk forward); the mental state of the lobotomized subject is not. As I said, this example, from Block, supports my claim against him that where the quale varies, so does the functional role. This being so, any system instantiating the Ramsey sentences of an adequate, true theory of M must also be said to have the same qualia as M, even if, like S, it is composed of a highly contrived organization of a bil-
CHINA AS ONE MIND

lion citizens of China.

Notice, by the way, that I have not begged the question against Block. Block intends his counterexample to show that system $S$ (composed of humans) would be psychofunctionally equivalent to $M$, without $S$ having qualitative mental states. To assume that $S$ has those states nonetheless would be to beg the question. But I have not done that. Instead, I have shown how Block’s own requirements for psychofunctional equivalence, considered from a Dennettian perspective, lead to a conclusion contrary to Block’s. Crucially, I do not assume, but demonstrate (using Block’s own example of lobotomized patients), that qualitative states would not play the same functional role were they qualitatively different. Block assumes that qualitative differences do not yield functional differences. I have not simply supposed that assumption false; I have shown it to be false.

4. CONCLUSION

I do not pretend to have shown psychofunctionalism to be immune to all objections. I have shown how to defend it against one counterexample, that of system $S$, the population of China as functionally equivalent to one mind, $M$. This example is of course representative of a whole family of counterexamples, but they do not exhaust the entire realm of counterexamples.

Nor do I pretend to have shown that a system $S$ really does have qualia simply by virtue of its functional equivalence to a system $M$ known to have qualia. I have shown only that, because of the nature of functional equivalence, we must regard $S$ as having qualia. But this is sufficient to meet Block’s challenge, since his claim was only that $S$, as described, does not seem to have the mental states in question. When viewed properly, it does seem to have them after all.

2. Ibid., 269, 273-74.

3. Ibid., 270 spells out a Ramsey functional correlate for the state of being in pain. See also 275.

4. Ibid., 271-72.

5. See Ibid., 301.

6. Ibid., 307.


10. Ibid., 8.

11. Ibid., 6-7.

12. Ibid., 268, 285.

13. Ibid., 269-71, 284-85. Of course, the reader will note that self-consciousness must be more than is needed to establish the presence of such lowly qualia as pain—after all, there are animals we do think to experience pain, etc., but to which we do not attribute consciousness of a self. But in this essay S is a system psychofunctionally equivalent to M, and M is a human mind. Therefore the personal stance is the stance appropriate toward M and S. The personal stance carries with it attribution of self-consciousness, whether or not it is needed for attribution of pain.

I do believe that non-self-conscious animals also have qualia as pain. I also believe that some sub-personal stance intermediate between the merely intentional and the fully personal can be justified. However, conciseness demands that I neglect this refinement in this essay.

14. Dennett does refer to an example (unpublished) similar to Block's, and refers to Block's article in a footnote, concerning the point at issue here (Op. cit., 153, 325n3), but does not take the present approach in rebutting this sort of counterexample. Dennett does
suggestion this approach, in his claim that even in one's own case one
decides questions of conscious experience by a "consideration of facts
about our current capacities and past activities, and the best theory
that can account for these." (Op. cit., 172) Still, Dennett's approach
to the whole problem of an "inner life", "consciousness", etc., for any
instantiation of a functionalist theory of mind is rather to sketch out
such a theory on a sub-personal level--a theory appropriate to the de-
sign stance (or intentional stance at best) (Op. cit., 153-164), and
then to offer the hypothesis "that you are a realization of this flow
chart, and that it is in virtue of this fact" that it seems both to you
and to others that you have consciousness, an inner life, qualia. (Op.
cit., 165) We are then challenged to deny this hypothesis in some
principled way. This appears to beg the question, since the exercise is
inspired (at least in part) by the counterexample of system $S$ which
presumably realizes whatever flow charts of theories $M$ realizes, but
which does not seem to possess qualitative mental states.

Block (op. cit., 289-90) discusses an inversion of this case: what
if you, a presumptive $M$, turned out to be a systematically organized
collection of little sentient beings? Wouldn't you still, as $M$, know
yourself ($M$) to have qualia? But anyone proposing Block's counterex-
ample does so, at least implicitly, denying that his own mind is com-
posed of ingredients similar to those of $S$. In other words, to offer
this challenge is to beg the question.