The Knowledge Intuition and the Ability Hypothesis

HUIMING REN *Zhejiang University of China*

ABSTRACT: I argue that the Ability Hypothesis cannot really accommodate the knowledge intuition that drives the knowledge argument and therefore fails to defend physicalism. When the thought experiment is run with, instead of Mary, an advanced robot Rosemary, for whom there presumably is no distinction between knowledge-how and knowledge-that, proponents of the Ability Hypothesis would have to give a far-fetched and counterintuitive explanation of why Rosemary wouldn't learn anything new upon release.

RÉSUMÉ : Je soutiens que l'hypothèse de la capacité (Ability Hypothesis) est incompatible avec l'intuition de connaissance qui sous-tend l'argument de la connaissance, et par suite échoue à valider le physicalisme. Si on opère l'expérience de pensée en remplaçant Mary par Rosemary, un robot perfectionné, dont on présume qu'elle ne peut distinguer le «savoir-comment» et le «savoir-que», les avocats de l'hypothèse de la capacité devraient élaborer une explication contre-intuitive et exagérément complexe des raisons pour lesquelles Rosemary n'apprendrait rien de nouveau après sa sortie de la chambre.

One of the main arguments against physicalism is the knowledge argument (hereafter, KA). This argument is grounded by a thought experiment conducted by Jackson:

> Mary is a brilliant scientist who is … forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there

***Dialogue* 51 (2012), 313–326.**
© Canadian Philosophical Association/Association canadienne de philosophie 2012
doi:10.1017/S0012217312000467

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

> is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like "red", "blue" and so on….
>
> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo,* there is more to have than that, and physicalism is false. (1982, p. 130)

The critical intuition that drives KA is: Mary learns *something new* after she is released from the black and white room and sees red for the first time. Specifically, she learns *what it is like* to see red. Let's call this intuition "the knowledge intuition." Given that the knowledge intuition is critical to KA, one way to defend physicalism would be to claim that the knowledge intuition simply is wrong: Mary has already known what it is like to see red before the release. But it seems that most people, including most physicalists, agree that the knowledge intuition is correct. Of course, these physicalists would then be obliged to explain the knowledge intuition in a physicalistically respectable way.

The Ability Hypothesis (hereafter, AH) proposed by Lewis (1990) and Nemirow (1990) is one of the main theories that have been advocated to fulfill this obligation. The central claim of AH is: knowing what it is like to see red is nothing but knowing how to IRR (imagine, remember, and recognize) an experience of red.[1] Given this claim, AH can defend physicalism while at the same time accommodating the knowledge intuition. The knowledge intuition is accommodated because AH agrees that Mary comes to learn something new upon release. Physicalism is defended because AH denies that Mary comes to learn any new factual knowledge upon release.

Doubts have been raised against AH;[2] in this article, I would like to raise another. My doubt is different from others' because it isn't directly aimed at the central claim of AH. Instead, I doubt that AH *really gets our intuition right*. But arguing about intuitions could be tricky, so I would like to explain my strategy before getting into the real business. To do this, I first need give an account of how I understand AH to work.

I think that the way AH manages to respond to KA is as follows: first, it twists the knowledge intuition a bit by claiming that the new knowledge we intuit to be acquired by post-release Mary is knowledge-how instead of knowledge-that and second, denies that upon release, Mary acquires any new factual knowledge by applying the distinction between knowledge-how and knowledge-that. Accordingly, my strategy is the following: I would question whether there really is a fundamental distinction between knowledge-how and knowledge-that. If there isn't such a distinction, or if at least it isn't relevant to the issue at hand, as I will try to show, it would seem that the knowledge-how twist is unmotivated. As a consequence, it would seem that AH cannot really accommodate the knowledge intuition.

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

It should be noticed that AH, as an account of phenomenal knowledge, may be correct; that is, knowing what it is like to have a certain experience may well be nothing but knowing how to IRR an experience of the relevant kind. But it is also important to notice that AH is *proposed to explain* the knowledge intuition—AH should not be taken to dictate what intuition we have. If we claim that since knowing what it is like to see red is nothing apart from knowing how to IRR an experience of red, the knowledge intuition underlying KA is only that Mary acquires new abilities to IRR an experience of red, we just put the cart in front of the horse. So, if the distinction between knowledge-how and knowledge-that won't make the key difference, and thus the knowledge-how twist on the knowledge intuition is unmotivated, to defend physicalism, proponents of AH would have to claim that the knowledge intuition is wrong. Then, some extra effort needs to be made to explain away rather than explain the knowledge intuition. But more importantly, AH would seem ill-motivated.

It might be helpful to take a look at Jackson's (2003) new position on KA here. Jackson agrees with proponents of AH that post-release Mary acquires new abilities, but he doesn't accept the knowledge intuition. Jackson makes a distinction between intentional properties and instantiated properties and argues that the property of what it is like to see red is an intentional property of an experience of red instead of an instantiated one. So he claims that Mary has already known what it is like to see red before the release—she can deduce it in the black and white room; and we have the wrong intuition because we fail to realize that qualia are representational, that is, we mistake an intentional property, namely, the property of what it is like to see red, for an instantiated one. According to Jackson, Mary's acquisition of new abilities is merely a consequence of her having an experience of red. This new position of Jackson not only in no way indicates that AH is right, but instead shows that AH is ill-motivated.

I will proceed as follows. In section 1, I will consider the distinction between knowledge-how and knowledge-that concerning an advanced robot. In section 2, I will investigate what proponents of AH have to say if we run the thought experiment grounding KA with this robot. In section 3, I will try to show why the best possible response from proponents of AH to their predicament presented by the thought experiment with the robot is defective.

1. The Distinction between Knowledge-How and Knowledge-That

Suppose that you want to let a robot know how to ride a bicycle. This robot is a very advanced one built in a future world. Let's call it "Rosemary." Let's further assume that, in the future world, there are no longer bicycles. Rosemary doesn't know how to ride a bicycle, nor does the designer of Rosemary know anything about bicycle or bicycle-riding. Rosemary hasn't ridden a bicycle yet. Since you are an expert on coaching people on how to ride a bicycle, it is your task to teach Rosemary how to ride a bicycle. However, though you are an expert, surprisingly you don't *know how* to ride a bicycle.[3] Now, imagine how you are going to coach Rosemary.

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

It seems that all you need to do is tell Rosemary what you know about bicycle-riding. It also seems that after Rosemary knows what you know about bicycle-riding, which is presumably knowledge-that, she would be able to ride a bicycle immediately—no practice at all is needed for her to know how to ride a bicycle. If that is the case, then, for Rosemary, there doesn't seem to be a distinction between knowledge-how and knowledge-that.

People may feel uneasy about this claim. For though much progress has been made in AI, e.g., building a computer that can beat the best chess player in the world, it is also notorious that the classical robotics meets a much greater difficulty in replicating or simulating even the simple human motion of grasping an object. There might be two broad questions about if a robot could acquire motor abilities that are close to those possessed by humans. One concerns whether we can replicate the mechanical properties of a human body. The other is with the enormous computational processing concerning performing an action. But both questions, it seems to me, present no difficulties in principle.

The more important and also more penetrating worry is with the distinction between knowledge-how and knowledge-that. That is, the claim that Rosemary would be able to ride a bicycle immediately after acquiring knowledge-that about bicycle-riding seems to most people contentious. So in the following, I will focus on addressing this worry.

The distinction between knowledge-how and knowledge-that, in cognitive science, is sometimes characterized in terms of the medium of representation. For example, Churchland says:

> Any competent golfer has a detailed representation (perhaps in his cerebellum, perhaps in his motor cortex) of a golf swing. It is a *motor* representation, and it constitutes his "knowing how" to *execute* a proper swing. The same golfer will also have a discursive representation of a golf swing (perhaps in his language cortex, or in the neighboring temporal and parietal regions), which allows him to describe a golf swing or draw it on paper. The motor and the discursive representations are quite distinct. … In one medium of representation, his representational achievements on the topic may be complete; while in another medium of representation, he has nothing. (1989, pp. 68-9)

The distinction between motor representation and discursive representation is supposed to explain why, for a human being with complete discursive knowledge about a certain action, practice is still indispensable for acquiring motor representation of the action, which is required for performing the action effectively.

Characterized in this way, the distinction might not apply to the knowledge representations possessed by Rosemary, for Rosemary might not have those features, whatever they are, that account for the different media of representation in human beings. There might be only one medium of representation for

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

Rosemary. Even if Rosemary also had to have two types of representation one of which is in control of performing actions, like a human being, for her, the two types of representation might not be *distinct*. For Rosemary, it might be the case that one type of representation could be translated into the other at will as the translation is sustained by a special module or mechanism. According to my conception, which I hope is not highly problematic, an *advanced* robot can complete any new task that requires high precision because every movement of its parts, such as eyes, limbs, fingers, etc., is precisely controlled by its brain. The orders from its brain will always be followed by its body, precisely. As an expert on bicycle and bicycle-riding, you cannot ride a bicycle because you cannot effectively control your body in a bicycle-riding way. Though you know exactly what you need to do when riding a bicycle, your body cannot follow the order from your brain (through your discursive representation of bicycle-riding) without an adequate motor representation; and that is the reason why you are unable to ride a bicycle. But if the body of an advanced robot is not under control of its brain, we would say that it is out of order, i.e., it malfunctions. It should be noticed that by claiming that the body of an advanced robot is always under full control of its brain when it functions well, I don't mean that the robot can do whatever it wills. What a robot can do would certainly be restrained by its material constitution, which realizes the capacities that the robot possesses. For example, a robot without legs certainly won't have the capacity of riding a bicycle. What I mean here is that, given that the robot possesses the capacity, say, riding a bicycle, it can acquire knowledge-how to ride a bicycle without practice.

So, if the distinction between knowledge-how and knowledge-that consists in different media of representation, the distinction seems not a principled one; that is, for other kinds of systems or creatures, there might not be such a distinction.

But one might think that the core of the distinction between knowledge-how and knowledge-that is not that they are different media of representation, but that they are different styles of representation: one tacit, the other explicit, in Dennett's (1982) terms. For example, the tacit knowledge-how in a Turing machine is the information stored in the read/write head instead of the tape, where the explicit knowledge representations are stored. Tacit knowledge-how is somehow hard-wired into the system. "That is, the read/write head is so physically constructed that it behaves as if it were following a set of primitive computational instructions. Information embodied in these primitive operations is neither explicit, nor potentially explicit, but tacit (O'Brien and Opie, 1999, p. 14)." So, the physical construct of a system has already carried some information that can be formulated as a set of instructions, and these instructions would be regarded as constituting tacit knowledge-how for the system. Knowledge-how in this sense might be deemed as distinct from knowledge-that *in principle,* as it is not explicit but tacit and could not be made explicit by the system itself.

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

However, knowledge-how understood as tacit knowledge in Dennett's sense would exclude procedural knowledge like knowing how to ride a bicycle because this procedural knowledge apparently is not *hard-wired* into a system. Though we sometimes call procedural knowledge like knowing how to ride a bicycle "tacit knowledge," in Dennett's terms, they should be called "potentially explicit knowledge." It might be a tough task for one who is able to ride a bicycle to write a manual for bicycle-riding, but the difficulty isn't a principled one, for there *is* a manual for bicycle-riding in principle. If we agree that there is a good sense in which knowing how to ride a bicycle could be rightly claimed to be knowledge-how, then even if the distinction between knowledge-how *as tacit knowledge* and knowledge-that is a principled one, it might not be relevant here. Of course, if knowing how to IRR an experience is tacit knowledge-how, we would be in trouble. But it seems implausible that knowing how to IRR an experience is hard-wired into us. Indeed, even proponents of AH agree that post-release Mary *learns* how to IRR an experience. So knowledge-how as tacit knowledge should not worry us here. From now on, knowledge-how in this article would not be understood as tacit knowledge in Dennett's sense.

Before we can move on, we have to remove another worry. The worry is that, though it might be true that Rosemary doesn't have to practice in order to know how to ride a bicycle, it still isn't clear whether she could know how to IRR an experience if she has all knowledge-that about IRR-ing an experience. In any case, how to IRR an experience is a mental action instead of a physical one.

This worry seems inappropriate. The consideration of the case of riding a bicycle about Rosemary is to establish that for Rosemary, there might not be any distinction between knowledge-how, be it knowing how to ride a bicycle or knowing how to IRR an experience, and knowledge-that. Very plausibly there is some, perhaps important, difference between knowing how to IRR an experience and knowing how to ride a bicycle as the former is a mental action while the latter is a physical one, but unless we can show that the difference makes a difference to the distinction between knowledge-how and knowledge-that, it is beside the point. When proponents of AH claim that knowing what it is like to have an experience is knowing how to IRR the experience, they don't claim that knowing how to perform a mental action is different *in kind* from knowing how to perform a physical one.

Though there might be other concerns that I am unable to address here, I hope that the above discussion at least makes it *prima facie* plausible that for Rosemary, if she has all the knowledge-that about IRR-ing an experience, she would be able to IRR the experience immediately.

2. The Thought Experiment with Rosemary

Now, the question is: what if the thought experiment grounding KA is run with Rosemary? Would she learn anything new after being released?

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

One might think that since Rosemary is a robot and therefore could not have any phenomenal feel, the thought experiment cannot be run with Rosemary in the sense that the thought experiment will not be related to conscious experience at all and therefore have no bearings on issues about experience.

Here a dialectical clarification is needed. In this article, some assumptions must be made, but since the target of the criticism is AH, as long as these assumptions are acceptable to proponents of AH, they won't weaken the criticism. Proponents of AH haven't said anything about whether robots, or creatures with different physical constitution from us humans, say, silicon-based instead of carbon-based, could have experience. The assumption that robots could have experience doesn't seem to conflict with the spirit of physicalism at all, so unless proponents of AH provide specific reasons for doubting this assumption, the above response would be inappropriate.

Under the assumption that Rosemary could have experience, once she is released from the black and white room, will we intuit that she learns something new? This is a hard question. But let us first suppose that we would answer "yes," as in the case of Mary. This answer immediately presents a problem for AH. Since, by assumption, Rosemary knows everything there is to know about experiencing red, she of course possesses full knowledge-that about IRR-ing an experience of red. Now, what prevents Rosemary from knowing how to IRR an experience of red before her release given that for her, there is no distinction between knowledge-how and knowledge-that? But if she already knows how to IRR an experience of red, she wouldn't acquire any new knowledge upon release *unless* knowledge of what it is like to see red is *not* knowledge-how to IRR an experience of red.

Notice that the case is different with Mary. Mary certainly also has possessed knowledge-that about IRR-ing an experience of red before her release, but she may not therefore know how to do that. Mary is a human being, the distinction between knowledge-how and knowledge-that applies to her, so knowledge-how to IRR an experience of red may still be new knowledge to her even though she has possessed knowledge-that about IRR-ing an experience of red. But the distinction doesn't apply to Rosemary. Once Rosemary possesses knowledge-that about IRR-ing an experience of red, she instantly acquires knowledge-how to do that.

Proponents of AH might claim that there isn't a way of IRR-ing an experience of red, or that one cannot possess full knowledge-that about IRR-ing an experience of red. But this would be a desperate move because it would mean that there is something over and above the physical facts or something that is beyond our cognitive capacities. So I don't think that proponents of AH will take this way out.[4]

Proponents of AH would have to say that Rosemary doesn't learn anything new at all. To take this option, they will have to either claim that we would intuit that Rosemary doesn't learn anything new upon release, or claim that we would intuit that Rosemary learns something new upon release, but the intuition is wrong.

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

What would we intuit when the thought experiment is run with Rosemary? The answer may not be so clear. But it is clear, and also accepted by proponents of AH, that we intuit, in the original thought experiment with Mary, that Mary learns something new upon release, be it new knowledge-how or knowledge-that. If, as proponents of AH might want to claim, our intuition in the thought experiment with Rosemary is that she doesn't learn anything new upon release, some difference between Mary and Rosemary must be held accountable for the differing intuitions. There are many differences between Mary and Rosemary, but it seems that given that Rosemary could have experience, only one difference is relevant here, namely, that there is a distinction between knowledge-how and knowledge-that for Mary while there is none for Rosemary. However, our intuition that Mary learns something new doesn't seem to have anything to do with the fact that the distinction between knowledge-how and knowledge-that applies to her. Rather, it is based on another intuition, namely, that under normal circumstances, one knows what it is like to see red only if one has experienced red. If in the thought experiment with Rosemary, we still have this base intuition, we would then consequently intuit that Rosemary learns something new. Further, it seems that if this base intuition is right with Mary, it is right with Rosemary, and if it is wrong with Rosemary, it is wrong with Mary. Indeed, if we have never heard of AH, the difference between Mary and Rosemary seems hardly a factor that would be considered when we intuit about conscious experience. However, proponents of AH certainly cannot accept that the base intuition is indifferent to Mary and Rosemary, for if they claim that this intuition is right with Rosemary, AH would not be a correct account of phenomenal knowledge; if they claim that this intuition is wrong with Mary, AH could not accommodate the knowledge intuition. To avoid this result, proponents of AH have to claim that the intuition is right with Mary but wrong with Rosemary.

Then an explanation is needed. What kind of explanation could proponents of AH provide? They could say, contrary to what it seems, the difference between Mary and Rosemary that is relevant here actually is the key: it explains why our intuition about Mary, a human being, is correct while our intuition about Rosemary, a robot, is not. When we try to figure out what happens to Mary after she leaves the black and white room and sees red for the first time, the story goes, we just put ourselves in Mary's shoes and see what would happen to us. Since Mary is a human being, as we are, our intuition could reliably indicate what would happen to her. However, when we put ourselves in Rosemary's shoes, our intuition would not reliably indicate what would happen to her, since she is a robot, she is *unlike us*. When we try to figure out what would happen to us if we are put in a certain situation, some special features of us might step in and play a role in our coming to a certain intuition. So, only if we put ourselves in the shoes of someone who shares the special features with us, could our intuition be a reliable indicator. Here, of course, the relevant special features of us are those that explains why the distinction between knowledge-how and knowledge-that applies to us.

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

This seems to be the only way to bring the relevant difference between Mary and Rosemary into play to explain why our intuition about Mary is right while the same intuition about Rosemary is wrong. One cannot say that our intuition is wrong simply because knowing what it is like to see red is nothing but knowing how to IRR an experience of red, as this blatantly begs the question. On the other hand, when we consider the thought experiment with Rosemary, we may clearly realize that, for her there is no distinction between knowledge-how and knowledge-that, so it would not be correct to say that our intuition is wrong since we miss this important feature of Rosemary.

However, the above explanation is unsatisfying. First, it would imply that we are restrained from coming to the right intuition by some special features of us, even though we have given the situation a totally rational assessment. Second, even if we concede that we might be somehow restrained, in my view, the best possible response from proponents of AH on how the special features that make the distinction between knowledge-how and knowledge-that apply to human beings could affect the way we come to a certain intuition about conscious experience is flawed. Let me elaborate.

3. The Possible Response from Proponents of AH and Its Flaw

Though AH, as an account of phenomenal knowledge, seems implausible to some people, including some physicalists, it does have its own intuitive appeal. On the one hand, we have the following strongly held intuition:

(i) Under normal circumstances, one knows what it is like to see red only if one has experienced red.[5]

On the other, it seems that:

(ii) Under normal circumstances, one knows how to IRR an experience of red only if one has experienced red.

Once noticing the parallelism between (i) and (ii), one would naturally hypothesize that knowing what it is like to see red is just knowing how to IRR an experience of red, that is, one would be naturally led to use (ii) to interpret (i). AH seems to be a result of this natural move.

In accord with this interpretation, proponents of AH might give the following explanation of why the special features that make the distinction between knowledge-how and knowledge-that apply to human beings are even relevant to the reason why we hold (i):

> We intuit that we come to learn something *new* when seeing colour for the first time *only because* we acquire *new* abilities when doing so. And we acquire new abilities when seeing colour for the first time *because* having an experience is the *only*

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

> practical way to acquire them. And having an experience is the only practical way to acquire them *because* we have those special features that make the distinction between knowledge-how and knowledge-that apply to us. If we didn't have those features, we wouldn't acquire any new abilities after seeing colour for the first time. And as a consequence, we wouldn't intuit that we learn anything new when doing so. (i) wouldn't be correct by its own right or under other interpretations. If we fail to realize this, then when we put ourselves in Rosemary's shoes, we might come to intuit that Rosemary comes to learn something new upon release *based on (i)*. But this intuition is wrong.

This response seems to be the best one proponents of AH could offer. On the one hand, it explains how we come to (i); on the other, in so explaining, it shows why only under the interpretation of (ii), would (i) be correct.

The following passage from Lewis could be viewed as textual support of the above response:

> Having an experience is surely one good way, and surely the only practical way, of coming to know what that experience is like. Can we say, flatly, that it is the only *possible* way? Probably not. There is change that takes place in you when you have the experience and thereby come to know what it is like. Perhaps the exact same change could in principle be produced in you by precise neurosurgery, very far beyond the limits of present-day technique. Or it could possibly be produced in you by magic. If we ignore the laws of nature, which are after all contingent, then there is no necessary connection between cause and effect: anything could cause anything. For instance, the casting of a spell could do to you exactly what your first smell of skunk would do. (1990, p. 500)

Lewis is making two points here. One is that having an experience is the only practical way of coming to know what that experience is like. The other is that having an experience is *not necessarily* required for acquiring phenomenal knowledge—by super neurosurgery or magic one can also come to know what it is like to have an experience. The first point could be taken as explaining why we come to hold (i): it is because having an experience is the *only* practical way of acquiring phenomenal knowledge. The second point, on the other hand, explains why having an experience is the only practical way of acquiring phenomenal knowledge: it is because we, or some features of us, are governed by some contingent causal laws. If we don't have those features, and therefore are free from those causal laws, having an experience would fail to be the only practical way of acquiring phenomenal knowledge, and consequently, we may not come to hold (i). If we still do, (i) would not be correct.

It seems that here Lewis presumes that phenomenal knowledge is knowledge-how. For if he doesn't, it is hard to see why some contingent causal laws are even related to the reason why having an experience is the only practical way of acquiring phenomenal knowledge. According to the standard reading of this

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

passage, Lewis is claiming that by other means, we also could put one in the *dispositional brain states* that are sufficient for phenomenal knowledge.[6] This reading is reasonable only if phenomenal knowledge is taken to be knowledge-how, or something close to it. Once we presume phenomenal knowledge consists in abilities to IRR experiences, and thus see that having experiences is the only practical way of acquiring these abilities, we would have a hint on how we come to hold (i): if by having an experience for the first time, we acquire some new abilities, and having an experience is the only practical way of acquiring these abilities, it is probably because we acquire some new abilities when we have an experience for the first time that we come to hold (i). This reason becomes more probable once we think that phenomenal knowledge is knowledge-how. Indeed, if knowing what it is like to have an experience is knowledge-how, and having an experience is the only practical way to acquire this knowledge-how, then (i) would just follow.

So far, I have done my best to formulate a possible response on the behalf of proponents of AH. According to this response, (i) concerning Rosemary would not be correct since Rosemary doesn't share with us those special features that are governed by contingent causal laws which make having an experience the only practical way of acquiring phenomenal knowledge, which in turn explains why we come to hold (i).

The core of this response is the proposal about the way we come to hold (i)—it is because we acquire *new abilities* when having an experience for the first time that we hold (i). But as we have seen, when proposing how we come to hold (i), proponents of AH seem to have already presumed their account of phenomenal knowledge. The point is not to find fault, for surely when one makes a proposal about a certain subject matter, one's other views on the subject matter would be relevant. In the case at hand, AH and the proposal on how we come to hold (i) are reciprocal: AH, as an account of phenomenal knowledge, provides a hint for the proposal; the proposal, if correct, in return would make AH a much more plausible account.

However, given the reciprocal relation between AH and the proposal, if there are other independent reasons that plausibly would lead us to hold (i), the proposal would seem *ad hoc*. Consequently, the above response would be far-fetched. But are there other independent reasons? I think that there are. Consider the relation between what it is like to see red, be it physical or phenomenal, and an experience of red. The relation seems to be *constitutive*: no creatures can have an experience of red while there is nothing it is like for them to experience red. Given this constitutive relation, plus that phenomenal concepts seem inexpressible in a public language, it is natural for us to come to hold (i). When one is having an experience of red, it would seem to one that there is something it is like to have the experience, and this something is constitutive of the experience. When one tries to tell other people what the experience feels like, the best one can do is appeal to the similarity and difference between different phenomenal feels of the category that the experience falls

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

into, say, colour experience. But if the listener has never had any experience of the category, usually one would not be able to give even a dim clue about how the experience feels. Then one would say, "Go to see red by yourself, you will know how it feels." Since what it is like to have an experience is constitutive of the experience, having an experience is a good way of acquiring knowledge of it. And since phenomenal concepts are inexpressible in a public language, having an experience would be the only practical way of acquiring knowledge of what it is like to have the experience. It is not hard to see why we come to hold (i).

It should be emphasized that all I am claiming here is that there are other independent reasons that plausibly would lead us to hold (i). I am not saying that the intuition is correct, or that these reasons themselves are right. The relation between an experience and what it is like to have the experience might not be constitutive, and phenomenal concepts might be perfectly expressible in a public language. But it is obvious that it is our *intuitive* understanding of the relation that it is constitutive. (Indeed, if it is not, the hard problem of consciousness may never arise.) And intuitively speaking, phenomenal concepts are inexpressible in a public language.[7]

These reasons are obviously independent of the reason why having an experience is the only practical way of acquiring abilities to IRR an experience. So the above response from proponents of AH is *ad hoc*. Furthermore, now we see that this claim is very counterintuitive: we intuit that we come to learn something *new* when seeing colour for the first time *only because* we acquires *new* abilities when doing so. For according to this claim, the relation between having an experience and acquiring knowledge of what it is like to have the experience is *causal*. But given the constitutive relation between an experience and what it is like to have the experience, the relation between having an experience and acquiring knowledge of what it is like to have the experience must be more intimate than a causal relation. When having an experience, one may introspectively focus on the phenomenal feel, and acquire phenomenal knowledge of it. The relation between phenomenal feel and phenomenal knowledge is not causal unless we take introspection to be a causal process. But taking introspection to be a causal process doesn't seem to be quite right, and is, at the very least, counterintuitive.

If we do come to hold (i) because of reasons that are independent of the features that make having an experience the only practical way of acquiring abilities to IRR an experience, then we would intuit that Rosemary does acquire something new upon release, for those reasons apply to Rosemary too, intuitively speaking.

The last resort proponents of AH have is to claim that if we come to hold (i) because of other reasons, (i) would not be correct. But this means, if we do come to hold (i) because of other reasons, that AH cannot really accommodate the knowledge intuition in the original thought experiment about Mary, as the knowledge intuition concerning Mary is based on (i).

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

4. Conclusion

The thought experiment with Rosemary for whom there presumably is no distinction between knowledge-how and knowledge-that reveals the internal tension of AH. Proponents of AH have to claim that upon release, Rosemary doesn't learn anything new. But to sustain this claim, they have to give a far-fetched explanation of how we come to hold the intuition that under normal circumstances, one knows what it is like to have an experience only if one has had that experience.

Acknowledgements

This project is supported by the Fundamental Research Funds for the Central Universities, and by MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 11YJC720036).

Notes

1 It should be noted that some people may think that the correct reading of AH (at least of Nemirow's version of AH) is that knowing what it is like to see red is nothing but knowing how to IRR *red* rather than *an experience of red*, where red is construed as a gerrymandered public property. While I agree that there is such a reading that might indicate the difference between Lewis and Nemirow, as is noted by Lewis (1990) himself, I believe that my reading of AH is the more popular one. On the other hand, Nemirow (2007), when defending AH against various criticisms, also takes AH under this reading. What is more important here is that the main point of this article doesn't depend on which reading of AH is the right one.

2 E.g., see Conee (1994), Tye (2000). For the response from proponents of AH, see Nemirow (2007).

3 It is a well-known phenomenon that one can coach on how to F very well without knowing how to F him/herself.

4 Here one might object that even if we take experiences of red to be physical (neurological), we still can claim that Rosemary cannot know how to IRR an experience of red before the release since an experience of red has not yet been realized (stimulated) in her brain. However, this objection, it seems to me, is based on the assumption that for Rosemary, there is a distinction between knowledge-how and knowledge-that, which has been shown to be problematic. It should be clear that for Rosemary to possess knowledge-that about IRR-ing an experience of red, no experience of red is required to be realized (stimulated) in her brain.

5 I am not claiming that this intuition is correct, but that we do hold this intuition strongly.

6 Clearly, there is another reading of the passage. One could interpret Lewis' claim that "precise neurosurgery or magic could produce the *exact same change* in one" as "precise neurosurgery or magic could produce the exact same change in one *as having an experience of red*." According to this reading, Lewis is claiming that we can make one have an experience of red by directly stimulating one's brain instead of making one see a red object as in normal circumstances. If this is the right reading, we surely could respond: but then one is still having the experience. But most people

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press

think, and I agree, that Lewis is claiming that by precise neurosurgery or magic, we can produce the exact same change *as possessing phenomenal knowledge* that one usually comes to possess via having the relevant experience. That is to say, Lewis is making a claim about *how we come to know* what it is like to have an experience instead of how we can have an experience.

7 Recently, it has been challenged that phenomenal concepts are inexpressible in a public language, see Ball (2009). But this challenge itself, on the other hand, shows that indeed, many people do hold the intuition that phenomenal concepts are inexpressible in a public language.

References

Ball, Derek

2009 "There Are No Phenomenal Concepts." *Mind* 118: 935–962.

Churchland, Paul

1989 *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge: MIT press.

Conee, Earl

1994 "Phenomenal Knowledge." *Australasian Journal of Philosophy* 72, 2: 136–150.

Dennett, Daniel

1982 "Styles of mental representation." *Proceedings of the Aristotelian Society* 83: 213–226.

Jackson, Frank

1982 "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127–136.

2003 "Mind and Illusion." in *Minds and Persons*. Edited by Anthony O'Hear, pp. 251–271. Cambridge University Press.

Lewis, David

1990 "What Experience Teaches." in *Mind and Cognition: A Reader*. Edited by W. Lycan, pp. 499–519. Oxford: Blackwell.

Nemirow, Laurence

1990 "Physicalism and the Cognitive Role of Acquaintance." in *Mind and Cognition: A Reader*. Edited by W. Lycan, pp.490-9. Oxford: Blackwell.

2007 "So *This* is What it's Like: A Defense of the Ability Hypothesis." in *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Edited by T. Alter and S. Walter, pp. 32–51. Oxford: Oxford University.

O'Brien, Gerard and Jon Opie

1999 "A Connectionist Theory of Phenomenal Experience." *Behavioral and Brain Sciences* 22: 127–148.

Tye, Michael

2000 "Knowing What It Is Like: The Ability Hypothesis and the Knowledge Argument." in *Reality and Humean Supervenience*. Edited by G. Preyer and F. Siebert, pp. 223–237. Rowman and Littlefield.

https://doi.org/10.1017/S0012217312000467 Published online by Cambridge University Press