Commentary on Katalin Balog, "In defense of the phenomenal concept strategy"

István Aranyosi, Bilkent University, Ankara, Turkey. (istvanaranyosi@gmail.com)

I.

Kati Balog offers an ingenious and elaborate defense of the idea that the physicalist phenomenal concept strategy (PCS) is as good as it gets in actually offering an explanation of the explanatory gap, against a recent argument by David Chalmers, who argues that PCS is as good as it gets in *not* being able to explain the explanatory gap.

Chalmers' argument is that:

(C1) If $P\&{\sim}C$ is conceivable, then C is not physically explicable.
(C2) If $P\&{\sim}C$ is not conceivable, then C cannot explain our epistemic situation.
(CC) Therefore, either C is not physically explicable, or C cannot explain our epistemic situation,

where P refers to the totality of microphysical facts, and C refers to the fact that we possess "phenomenal concepts with the relevant key feature (e.g., being constituted by an instance of the referent) posited by a physicalist account of phenomenal concepts" (Balog, section III). "Our epistemic situation" refers to the fact that we observe or intuit an explanatory gap between physical and phenomenal facts –we can conceive P without the phenomenal facts.

The conclusion is that PCS cannot be *both* explanatory and physicalistic, as the disjunction expressed in the conclusion is exclusive.

II.

Balog argues that Chalmers' premises are not true without qualification. She argues that according to the physicalist, C can be conceptualized in two ways: both by using phenomenal language and by using physical language. We get, therefore, two conceptualizations of C: C_{Phen} (phenomenal) and C_{Phys} (physical). Now she admits that the following two conditionals are non-vacuously true on her view:

(B1) If $P\&{\sim}C_{Phen}$ is conceivable, then C_{Phen} is not physically explicable.
(B2) If $P\&{\sim}C_{Phys}$ is not conceivable, then C_{Phys} cannot explain our epistemic situation.

The conclusion we get now is:

(BC) C_{Phen} is not physically explicable, or C_{Phys} cannot explain our epistemic situation.

But this conclusion does not express an exclusive disjunction, and it is compatible with:

(BC*) C_{Phen} explains our epistemic situation & C_{Phys} is physically explicable.

Balog happily embraces (BC*) because (a) it is no threat to physicalism, and (b) it is explanatory. The only further proposition we need for (a) and (b) to be true is:

(*) C_{Phen} and C_{Phys} are conceptualizations of the same fact.

All we need, according to Balog, to answer Chalmers' argument is the conceivability of (BC*) and (*). The fact conceptualized by C_{Phen} and C_{Phys} is that we possess a phenomenal concept with certain features. So the fact that we possess a phenomenal concept with certain features is conceivably both physically explicable (under the physical conceptualization) and explanatory of our epistemic situation (under the phenomenal conceptualization).

III.

I agree with Balog that this reply to Chalmers' argument is effective, and that Chalmers' response to it is toothless, but I think it suffers from an internal weakness, potentially fatal to the whole project of Balog's.

Let me first quote Chalmers' reply (2007: 180-1) to a reasoning like Balog's above, in order to contrast it with my own reply to Balog:

> "On this account, even if there is a sort of explanation of the explanatory gap in terms of features of phenomenal concepts, **(1)** *the explanatory gap recurs just as strongly in the explanation of phenomenal concepts themselves*. Because of this, the strategy may make some progress in diagnosing the explanatory gap, but it will do little to deflate the gap.
>
> A proponent may suggest that just as the first-order explanatory gap can be explained in terms of second-order features of phenomenal concepts, **(2)** *the second-order explanatory gap concerning phenomenal concepts can be explained in terms of third-order features of our concepts of phenomenal concepts, and so on*. Alternatively, an opponent may suggest that **(3)** *the second-order explanatory gap can be explained in terms of the same second-order features of phenomenal concepts that explain the first-order explanatory gap*. The first move here obviously leads to a regress of explanation, and the second move leads to a circular explanation. (…)
>
> A proponent may also suggest that to require that the key psychological features be physically explicable is to set the bar too high. On this view, all that is needed is a psychological explanation of the epistemic gap that is compatible with the truth of physicalism, not one that is itself transparently explainable in physical terms. However, an opponent will now question the compatibility of the account with the truth of physicalism. **(4)** *Just as the original explanatory gap gave reason to think that consciousness is not wholly physical, the new explanatory gap gives reason to think that phenomenal concepts are not wholly physical.*" (my numbering and emphasis)

Let me address each of the four points that appear in the emphasized passages above.

(1): The fact that an explanatory gap occurs at a higher level, namely, between *P* and, adopting Balog's framework, C_{Phen} is no surprise for the PCStrategist at all. PCS contains at least two claims: that phenomenal concepts are not physical-functional concepts, and that they have some further features that are supposed to explain the first-order gap, i.e. that between P and some phenomenal property *Q*. Since it is part of PCS that phenomenal concepts are not physical-functional concepts, the PCStrategist is well aware that there is no physical explicability of phenomenal concepts; otherwise they would be physical-functional concepts.

(2) and (3): there is no need for third-order features of phenomenal concepts to explain the second-order gap. It is explained by the facts that: phenomenal concepts are not physical-functional concepts and that physical-functional concepts picking out facts in the actual world are a priori entailed P.

(4): The PCStrategists reply is quite obvious, and is the reverse of Chalmers' statement: Just as the original explanatory gap gave no reason to think that consciousness is not wholly

physical, the new explanatory gap gives no reason to think that phenomenal concepts are not wholly physical. Since the PCStrategist argues against the move from an epistemic gap between *P* and *Q* to an ontic gap between them via the idea of phenomenal concepts, she will do the same now with respect to P and C_{Phen}: there is an epistemic gap between *P* and C_{Phen} because of the latter being a phenomenal concept and not a physical one. If earlier it was accepted that the epistemic gap between *P* and *Q* does not entail an ontic gap between them because *Q* is grasped via a *phenomenal* concept rather than a physical-functional one, then it should be (or could be) admitted now, at a higher level, that the epistemic gap between P and C_{Phen} does not entail an ontic gap between them, because C_{Phen} is what it is – again, a *phenomenal* conceptualization rather than a physical-functional one. As Balog correctly points out, if there is circularity here, it is not vicious but explanatory. Finally, via Balog's premises (*) and (BC*), we get that PCS is not only explanatory but physicalistic too.

IV.

Let's, finally, turn to my argument against Balog's account. I also think, like Chalmers, that there is a new gap emerging from Balog's framework. If I'm right, this gap is both physicalistically inexplicable in Balog's framework and entails the falsity of physicalism, given one of Balog's own premises.

My proposed gap is between P and the fact *F* that there is a phenomenal conceptualization - C_{Phen} - of the fact – *C* - that there are phenomenal concepts with the relevant features Balog ascribes to them. I would like to stress that this fact is different from C_{Phen} itself, so the gap is not captured by:

(B1) It is conceivable that $P \,\&\sim C_{Phen}$

C_{Phen} expresses in phenomenal vocabulary the fact that there are phenomenal concepts that have some features. Let φ stand for the fact that there are phenomenal concepts that have some features. Let ψ stand for the fact that C_{Phen} expresses φ.

Suppose that P & ~ψ is not conceivable. Then the fact that φ is phenomenally expressed by C_{Phen} could have been deduced from *P*, which means that ψ does not distinguish us from conceivable zombies, and, therefore, does not explain our epistemic situation.

Suppose that P & ~ψ is conceivable. Then, following Balog's recipe above, this is only consistent with both physicalism and explanatoriness regarding our epistemic situation if there is a physical conceptualization of ψ. But ψ contains phenomenal vocabulary, and on the assumption of PCS that phenomenal and physical concepts are distinct and not a priori related, there can't be a physical conceptualization of ψ. That is, there can't be a physical conceptualization of a phenomenal conceptualization – that would destroy the phenomenality of that conceptualization. But given Balog's implicit assumption that according to physicalism there is a physical conceptualization of any actual fact, the impossibility of a physical conceptualization of ψ entails the falsity of physicalism.

So the conclusion of my argument is:

(AC) Either ψ is explanatory regarding our epistemic situation, or physicalism is false.

V.

It might be objected that I make a use/mention fallacy in the above argument. More precisely, there can be a physical conceptualization ψ_{Phys} of ψ, because either in ψ_{Phys} or in both ψ_{Phys} and ψ, C_{Phen} is mentioned rather than used.

I have two independent replies to this objection.

A.

(α) Suppose that C_{Phen} is mentioned in ψ. In order for ψ to be explanatory of our epistemic situation, it has to be the case that P & $\sim\psi$ is conceivable, even under the assumption that ψ has the phenomenal vocabulary only mentioned. But this means that from the point of view of whether ψ is explanatory, the use/mention issue does not make a difference. So in ψ, C_{Phen} behaves as if used, even if it is mentioned.

(β) Now suppose that C_{Phen} is mentioned in ψ_{Phys}. In order for ψ_{Phys} to be a priori entailed by P, and so to render ψ_{Phys} physically explicable because C_{Phen} is only mentioned in ψ_{Phys} , the behavior of C_{Phen} has to be different when mentioned from when used, otherwise the use/mention issue wouldn't make a difference.

But since ψ_{Phys} is merely a conceptualization of ψ, they have to express the same fact. If physicalism is true, they express the same *physical* fact. What we get, finally, is that whether it makes a difference or not whether the very same phenomenal vocabulary is used or mentioned in stating the very same physical facts does not supervene on the totality of physical facts. The PCStrategist cannot now appeal to the fact that phenomenal concepts are not physical-functional concepts; the argument concerns the *very same* phenomenal concepts, and indicates that their behavior does not supervene on physical facts. Since their behavior is no less a fact than any other fact, we get some fact that does not supervene on physical facts. So again the victim is physicalism itself.

B.

My second reply is more speculative in that it depends on the interpretation of certain features Balog ascribes to phenomenal concepts. Let me first quote Balog.

> "(…)In the case of phenomenal concepts, e.g., the concept PAIN, constitution matters for reference, both in terms of how the reference is determined, and in terms of how the concept cognitively presents its reference. More precisely, on this view, every concept *token* applied to *current* experience is constituted by *that token experience*, and this fact is crucial in determining the reference of the concept. Not only is it the case that a token state that realizes a token concept also is a token of the referent, but it is *because* the concept is so constituted that it so refers. There are, of course, applications of phenomenal concepts that are, on this theory, *not* constituted by occurrent phenomenal states; e.g., applications of phenomenal concepts to one's past or future experience, to other peoples' experiences, etc. But the canonical, first person, present tense applications are always so constituted and these other applications are derivative on first person applications."

I want to argue that, given what Balog says in the quotation above, mentioning phenomenal vocabulary *by us* (as opposed to conceivable zombies) implies using them. Suppose that what Balog says about non-canonical uses (applications) of phenomenal concepts is true, namely, that they are derivative of canonical uses. In canonical uses, when we apply a concept like, e.g., PAIN, we apply it, according to Balog, directly to the reference of the concept, because the reference constitutes the very application of that concept. Now what happens in non-canonical

applications of PAIN? Such an application is, for instance, when some pain is not currently experienced but remembered. If it is derivative of the canonical application, then it implicitly mentions that canonical application. `I remember being in pain yesterday` would go something like this `I bear relation R to a time in the past at which I applied PAIN to my *then current* experience`. We can formalize this statement using a quantified tense logic containing the two-dimensional operators N (`it is now the case that`) and T (for `it is then the case that`), besides the usual, one-dimensional P and F, for past and future (Ulrich Mayer 2008), respectively, as follows:

(a) N(*i* bear *R* to (PT (*i* apply PAIN to experience)))

But since the application of PAIN to the current pain one is undergoing constitutively involves that pain, we can replace T(*i* apply PAIN to experience) with T(*i* am in pain). So the formula becomes:

(b) N(*i* bear R to (PT (*i* am in pain)))

(b) uses PAIN rather than mentioning it. T is modified by P, so T refers to a past time at which I undergo pain; so it does not follow that I am in pain now.

But notice that the situation with ψ mentioning phenomenal concepts is the very same, the only difference being that ψ is not about the past application of phenomenal concepts, but about their application in general and at all times. In our language we formalize `at all times it is the case that *q*` as:

(c) (~P~*q*) & (~F~*q*) & N*q*

So whatever proposition *p* exactly asserts about PAIN by mentioning it, we can formalize this fact as:

(d) ~P~((PAIN is applied) → *p*)) & ~F~((PAIN is applied) → *p*)) & N((PAIN is applied) → *p*))

If it is true that PAIN is applied at some time, then:

(e) PT(PAIN is applied) or FT(PAIN is applied) or N(PAIN is applied)

which then implies:

(f) PT(someone is in pain) or FT(someone is in pain) or N(someone is in pain).

So the conclusion is that, given the features Balog attributes to phenomenal concepts via the constitutional view quoted above, it doesn't matter whether ψ uses or mentions phenomenal vocabulary. Mentioning it is no different from a non-canonical application of its component concepts, which, in turn, will use them, albeit derivatively. The derivative use of concepts, in general, is no less use than the canonical one, as witnesses by the following sentence:

(g) The word `cats` is used to express truths about cats.

This sentence is both about `cats` and about cats. The same is the case, I claim, with the content of ψ, if Balog's constitutional account is right.

References:

Balog, K. (2008) In Defense of the Phenomenal Concept Strategy. *Manuscript*.

Chalmers, D. J. (2007) Phenomenal Concepts and the Explanatory Gap. In T. Alter and S. Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.

Mayer, U. (2008) 'Now' and 'Then' in Tense Logic. *Journal of Philosophical Logic*, online first. http://www.springerlink.com/content/45210j50629h4422/