COGNITIVE ARCHITECTURE AND THE EPISTEMIC GAP:
Defending Physicalism without Phenomenal Concepts

Peter Fazekas

School of Philosophy, Psychology and Language Sciences
The University of Edinburgh

Institute for Philosophical Research
Hungarian Academy of Sciences

Abstract

The novel approach presented in this paper accounts for the occurrence of the epistemic gap and defends physicalism against anti-physicalist arguments without relying on so-called phenomenal concepts. Instead of concentrating on conceptual features, the focus is shifted to the special characteristics of experiences themselves. To this extent, the account provided is an alternative to the Phenomenal Concept Strategy. It is argued that certain sensory representations, as accessed by higher cognition, lack constituent structure. Unstructured representations could freely exchange their causal roles within a given system which entails their functional unanalysability. These features together with the encapsulated nature of low level complex processes giving rise to unstructured sensory representations readily explain those peculiarities of phenomenal consciousness which are usually taken to pose a serious problem for contemporary physicalism. I conclude that if those concepts which are related to the phenomenal character of conscious experience are special in any way, their characteristics are derivative of and can be accounted for in terms of the cognitive and representational features introduced in the present paper.

Keywords: unstructured representations, phenomenal consciousness, phenomenal concepts, epistemic gap

COGNITIVE ARCHITECTURE AND THE EPISTEMIC GAP:
Defending Physicalism without Phenomenal Concepts

1. Introduction

This paper is about phenomenal consciousness and cognitive architecture. In what follows I shall argue that acknowledging the epistemic gap between conscious experience and the corresponding physical processes does not force us to abandon physicalism. To this extent the present account is in agreement with the so-called *Phenomenal Concept Strategy* (cf. Loar, 1990; Papineau, 2002; Stoljar, 2005). However, the aim of this paper is not to defend Phenomenal Concept Strategy; neither is it to provide a new version of it. Instead, I shall offer an *alternative approach* to the very same conclusion—namely that physicalistic explanations can account for the occurrence of the epistemic gap.

The present account disagrees with Phenomenal Concept Strategy on the role of phenomenal concepts. I shall claim that by focusing at phenomenal *concepts*, the Strategy misidentifies the cognitive underpinnings of the epistemic gap. The peculiar characteristics of conscious experience (e.g. Chalmers, 1996, 2003; Jackson, 1982; Levine, 1983; Shoemaker, 1981) are due to certain cognitive features of the experiences themselves, and not to the concepts—phenomenal or other—built upon them. That is, the conceptual features that the Phenomenal Concept Strategy investigates are not the genuine causes of the special characteristics of phenomenal consciousness but only symptoms which can themselves be explained in terms of the features of sensory representational states.

2. Fodorian Framework

The chosen framework of this endeavour is Fodor's theory (Fodor, 1987, 1998, 2008). Note, however, that this paper does not argue for the Fodorian view; it simply utilises his theory of concepts. On Fodor's account sensory and phenomenal concepts are not in any way special. Such a framework is helpful given that what the paper intends to show is that in the context of phenomenal consciousness conceptual features are at best explananda but not explanans.

The paper utilises the following picture of cognitive/perceptual architecture. The sensory system which consists of transducers and pre-conscious, modular processes connects the cognitive system to the environment. Transducers convert the physical input (different forms of energy) into signals usable by the rest of the sensory system which processes the information transformed by transducers.

Sensory representations or experiences form the interface between the sensory system and the cognitive system—they are the outputs produced by the sensory system and the inputs consumed by the central cognitive/conceptual system (which in turn controls behaviour). It is the particular sensory representation that transfers information about a distal layout (the object represented) for the conceptual system.

Inside the conceptual faculty what one can find are concepts and files. Concepts are atomic Language of Thought symbols which are nomically locked to the properties they are concepts of. Concepts are associated with files which are placeholders for information about the property being represented by the concept. Files have two slots, one for perceptual prototypes (P-slot) and one for abstract knowledge (A-slot). Typically, P-slots get filled by perceptual templates via perceptually experiencing the object in question. The content of an A-slot, on the other hand, is what one reports when explaining the meaning of a particular concept. However, files are not constitutive of concepts—on the Fodorian view concepts are strictly atomic LOT symbols. Nevertheless, concepts might have complex representational content (as most of them do) which might be accessible for further cognitive processing via the associated files if those are not empty (Margolis, 1999).

Take the example of seeing a day-old chick. When one, upon seeing a day-old chick, recognises it and entertains the concept DAY-OLD CHICK what happens in one's conceptual faculty is the following. The concept DAY-OLD CHICK is an atomic Language of Thought symbol which is asymmetrically causally dependent on day-old chick occurrences. The file associated to the DAY-OLD CHICK concept has its P-slot filled with a perceptual template of a day-old chick, and its A-slot containing abstract knowledge of day-old chicks, like "it is an animal", "it has yellow feather" and so on.

Perceptual templates filling P-slots are typically, but not necessarily formed via perception—they might also be formed from abstract knowledge. This is what, following Fodor and Pylyshyn, might be called *simulation of look* in imagery (Pylyshyn, 2002, 2003). According to these authors, imagination is simulation based on concept deployment. Combining concepts with perceptual prototypes in the P-slot of their associated files are accompanied by combining these perceptual prototypes. This is how one might acquire complex perceptual prototypes from abstract knowledge.

For instance, consider an ornithologist, who, as it happens, has not yet seen any day-old chick so far—not even pictures thereof. However, she knows everything that can be learned from descriptions about the look of day-old chicks. This knowledge, together with her previous encounters with other birds via which she could form perceptual templates of wings, claws, bills etc. make her able to imagine day-old chicks. What she does is entertaining the combined concepts YELLOW, FEATHER, LEG and so on. Given that the P-slots of these concepts have already been filled, entertaining these concepts is accompanied by visual memories of yellow, feather, leg, etc. Roughly speaking, the visual image she constructs on entertaining DAY-OLD CHICK is one that arises from remembering actual perceptions of feathers, legs, the colour yellow, and so on.

3. Structured vs. Unstructured Perceptual Representations

Now compare what it is like to see a day-old chick with what it is like to see something red. Surely, there are major differences in the two experiential states. However, there is one particular difference which is of great importance. The experience of seeing a day-old chick could not be like what it is without *certain information about structure* being conveyed in the experience. So, for example, the experience of seeing a day-old chick has as its constituent part in it the experience of seeing two legs with appropriate shapes and being in a certain position relative to each other and to other parts.

Seeing a day-old chick is a complex experience that carries a lot of information about spatial structure that can be recovered by concepts. One could generate lengthy and relevant descriptions of what day-old chicks are like, on the basis of a day-old chick picture. The stimulus is complex, and the experience makes this complexity available to cognition.

Now compare this with the experience of seeing something red. Its stimulus is complex too—it is a type of surface reflectance—still the experience does not make this complexity available to cognition in any way; the experience of seeing something red has no structure in the above sense.

This leaves us with one of the fundamental claims of this paper: there are structured as well as *unstructured* representations in our cognitive architecture. Some might object here that my example of 'seeing something red' does not support the claim that colour experiences are unstructured. We do not see free-floating colours but coloured objects. Some even argue that we see the shape of an object in virtue of seeing its colours (Smith, 2010). However, the claim here is not that there is no structural information conveyed in everyday cases of seeing colours. Seeing a red tomato, for example, is a complex experience. The claim rather is that complex experiences are constituted by a structure of simpler experiences—one is able to discern constituent parts in one's complex experi-

ences,—and that there are certain constituents which themselves have no further constituent structure.

An experience possesses constituent structure if and only if (1) it has discernible aspects or attributes (discernible at least for trained subjects) which (2) can occur as experiences on their own as well as in other perceptual experiences. Constituent structure in most cases has the function of mapping some structure of the represented object—the pattern of relations among representational constituents conveys information about a certain external pattern of relations in the environment. So, for example, it is possible to discern the experience of seeing an appropriately shaped leg as a constituent part of the experience of seeing a day-old chick, and it is also possible to experience an appropriately shaped leg on its own without necessarily experiencing the day-old chick.

Shape experiences, in general, have constituent structure. It is possible to undergo visual experiences of certain discernible parts of virtually all shapes without experiencing other parts at the same time. This is also reflected in some of the most popular accounts of visual shape representation (e.g. Biederman, 1987; Marr, 1982).

On the other hand, some perceptual representations are unstructured. First, on the personal level, the what-it-is-likeness of being in these states possess no constituent structure in the above sense, second, on the sub-personal level, these perceptual representations are unstructured for further processes. These unstructured representations do they map the structure of the represented object. They only indicate the presence of the represented object—they are "inner light bulbs" which reliably go on whenever a complex state of affairs obtains in the environment. I shall call such unstructured sensory representations *monadic markers*. They are monadic since they possess no structure, and they are markers since they do not map the structure of the represented object, only indicate its presence.

Monadic markers typically do have discernible attributes, so-called *primitive sensory features*. However, primitive sensory features cannot occur as standalone experiences, and have no discernible experiential features other than themselves. Consider, for example, unique hue experiences (e.g. a Ganzfeld experience). A unique hue experience has hue, saturation and lightness as its primitive sensory features. One is able to discern these attributes of, say, a red experience, however, one is unable to undergo an experience of a particular hue without experiencing a certain amount of saturation and lightness at the same time. That is, unique hue experiences in general—and experiences of seeing something red (as e.g. a constituent of seeing a red tomato, see above) in particular—are monadic markers.

4. Encapsulation

So far, it has been argued that there are unstructured representations—monadic markers—mediating between sensory input systems and the central conceptual system. Monadic markers, however, are unstructured only for certain processors, namely those residing within higher cognition.

Consider, for example, the experience of seeing something red. It is a mediator between the low level processes of the visual input system and the high level processes of the central system. For higher cognition, this experience is unstructured, and characterised by its primitive sensory features (specific values of hue, saturation and lightness). On the other hand, from the perspective of early visual processes it is a rather structured state of the mind—it is a result of a quite complex set of mechanisms taking place within the visual input system.

However, no matter how complex and structured these low level processes are, they remain hidden from higher cognition. Mechanisms within the input systems are *modular* and *encapsulated*, at least to a degree which makes these processes inaccessible to conscious reflection. One might have detailed knowledge about what happens in one's visual system when one has the experience of seeing something red; nonetheless, one is unable to consciously reflect on actual processes of one's early visual system, nor can one consciously influence these processes in any way. For such an interaction to take place, a sufficiently direct and detailed information transfer would be required between these two levels of representation. This kind of information transfer certainly does not obtain between colour processing and propositional representation.

5. Causal Role Exchange and Functional Unanalysability

Now that the stage is set, I shall draw attention to a consequence of there being monadic markers in our cognitive architecture—playing a mediator role between the sensory systems and the central system—which turns out to be of crucial importance.

This consequence is the possibility of *causal role exchange*: monadic markers, and unstructured representations in general, are able to exchange the causal role they play in a system. The very same unstructured representation might play different causal roles in the same system at different times, and also, different unstructured representations might play the very same causal role in the same system (or in different systems of the same type).

Intuition pumps back up this claim. In what follows two cases shall be considered: the case of colour versus shape experience inversion, and the case of role exchange between conceptual atoms versus sentences.

First, consider the contrast between colour experience inversion scenarios and visual shape experience inversion scenarios. It is quite easy to conceive of a colour experience inversion scenario where ripe tomatoes look to colour-inverted subjects the way grass looks to the rest of the population without there being any further difference in their cognitive architecture.

Contrary to this, visual shape experience inversion scenarios are harder to conceive of. If, due to some optical distortion, a subject misperceives circles as squares this tends to give rise to mistaken inferences in her mind about the shapes of certain objects. A fairly complex change in subsequent processing is needed to straighten out all geometrical inferences related to squares and circles in the subjects mind; and this is only a very simple case of shape inversion.

Second, take two atomic symbols of Mentalese, X and Y. Suppose that, in subject A's mind, X is locked to spoons, and has an associated file containing a perceptual prototype of spoons and relevant abstract knowledge. Also in subject A's mind, Y is locked to knives and has an associated file with proper contents. However, in subject B's mind, X is locked to knives and is associated with the knife-file, whereas Y is locked to spoons and has the spoon file associated with it. If the Fodorian view of concepts (in which atomic concepts are locked to the properties they represent and are associated with relevant knowledge) is right, then the role switch just sketched seems straightforwardly possible.

Contrary to this, the semantic inversion of sentences (complex linguistic representations) is an utterly strange idea. Imagine that the sentence "Bochum is in Germany" expresses a geographical fact for subject A, but for subject B it expresses the very zoological fact "giraffes are taller than dogs" expresses for subject A.

The moral that follows is that complex representations are much more tightly embedded in a system than unstructured representations. Two complex representations cannot assume each others causal role, for instance, without a significant reorganisation of the whole system, whereas unstructured representations seem to be able to exchange their causal role freely.

Although causal role exchange between unstructured representations seems to be a coherent idea, it does not follow that in adult subjects' minds such an exchange can easily happen. For unstructured representations UR1 and UR2 to exchange causal roles all the causal connections UR1 has to other states and behaviour need to be assumed by UR2 and vice versa. That is, for minds in which a large number of learned connections are already firmly in place actual causal role exchange seems to be impossible. Therefore, what the possibility of causal role exchange really means is this: the role an unstructured representation actually plays in a cognitive system could have been filled by some other unstructured representation equally well, in other words, the unstructured representa-

tion in question does not fill the role it actually fills necessarily.

This is an important consequence of being unstructured since it entails the *functional unanalysability* of unstructured representations. Since an unstructured representation does not fill the role it actually fills necessarily, knowing what role it fills (no matter how detailed the description is) does not help in specifying what unstructured representation it is that actually fills the role. Although the causal role actually filled by the unstructured representation is characteristic of the unstructured representation as a part of the actual system, since the very same causal role could have been filled by a different unstructured representation, it does not characterise the actual filler *uniquely* and thus does not distinguish it from other possible fillers.

6. Explaining the Epistemic Gap

With all the resources at hand, it is time to see how the approach proposed here can account for the presence of the epistemic gap.

First, recall the case of the ornithologist who has never ever seen a day-old chick. Though the A-slot of the file associated to her DAY-OLD CHICK concept is filled with detailed propositional knowledge about day-old chicks, its P-slot is empty. Now consider that this is the first time she tries to imagine a day-old chick. Based on descriptions mentioning such simple shapes (e.g. feather, wing, claw, etc.) and colours (e.g. yellow) which the ornithologist has prior experience of she is able to simulate the look of a day-old chick. That is, her abstract knowledge drives her imagination resulting in a perceptual prototype which now fills the so far empty P-slot. Note that the P-slots of the files associated with the concepts YELLOW, FEATHER, and so on, must have been filled in order for abstract knowledge to be able to generate the day-old chick prototype by simulating the look.

Now compare this case with the case of Mary (Jackson, 1982, 1986) the future neuroscientist, who has never ever seen anything red. However, she knows everything there is to know—in terms of descriptions—about seeing something red. That is, just like in the case of the ornithologist, though the A-slot of the file associated to Mary's concept RED is filled with detailed propositional knowledge, its P-slot is empty. Nonetheless, there is an important difference between Mary and the ornithologist, namely that whereas the ornithologist is able to fill the relevant P-slot with a day-old chick prototype by simulating its look based on constituent structure, Mary (before leaving her room) is unable to fill the P-slot of the file associated to the concept RED with a red experience prototype. What makes the difference here is the fact that while the experience of seeing a day-old

7

chick is structured the experience of seeing something red is not. The latter is a monadic marker, an unstructured representation, and as such its perceptual prototype cannot be simulated on the basis of the abstract knowledge Mary could acquire within her chamber.

To see why this is so, let's first unpack what happens with Mary. Before her release, she learns from books that seeing something red is identical with a salient neurological response pattern in subjects' V4 what she calls "neuro response X", and so forms the concept NEURO RESPONSE X of it. That is, the A-slot of the file associated to pre-release Mary's NEURO RESPONSE X concept is filled with relevant neurological information, but its P-slot is empty—she has no idea of the relevant phenomenal character of occurrent neuro responses.

Then, after her release, Mary is shown a piece of paper with a patch of red on it (Nida-Rümelin, 1996). She locks the dummy concept XYZ to this kind of stimulus. (According to Fodor's innateness thesis, our conceptual system is full of dummy concepts waiting to get locked to a new kind of stimulus. One does not need to buy innateness here though, for one could argue that Mary acquires a new concept XYZ which is nomically locked to perceived redness.) Whereas the P-slot of the associated file is filled with the perceptual prototype of red, its A-slot is at least nearly empty. Nonetheless, deployment of the concept XYZ helps Mary in recognising and categorising new varieties of the red stimulus, and in this way, contextual information can fill the A-slot of the associated file with information like "it is the colour of London buses", or "it is the colour of ripe tomatoes".

Note, that there is no a priori connection between Mary's XYZ and NEURO RESPONSE X concepts. First, Mary cannot fill the A-slot of XYZ with information like "neuronal activation X" based on introspection when seeing something red, because this level of processing is encapsulated—inaccessible to conscious reflection.

Second, Mary is unable to fill the P-slot of NEURO RESPONSE X up with a perceptual template of a red experience solely on the basis of the content of its A-slot. Since it is a monadic marker, an unstructured representation, it has no constituent structure that could help Mary (like prior experiences of wings and claws helped the ornithologist). Nor has Mary access to the primitive sensory features of red experiences—primitive sensory features do not form standalone experiences, the only way to experience them is as features of the overall experience (seeing something red), which is exactly what Mary lacks.

Moreover, neither structural nor functional information conveyed by the abstract knowledge in the A-slot is of any use. Structural information cannot help in simulating monadic markers due to the fact that monadic markers lack any structure, and functional information cannot help either due

to the functional unanalysability of monadic markers.

The only way for Mary to connect the two concepts is via filling the A-slot of the XYZ concept with contextual information she is familiar with from her pre-release studies. If, for example, Mary learns that XYZ is utilised when seeing the colour of London buses, and if she has learnt inside her room that NEURO RESPONSE X answers to the colour of London buses in neuro-typical individuals, then she can conclude that XYZ is NEURO RESPONSE X that is, that these two concepts co-refer. In effect, the A-slots of the two concepts can get merged—all the conceptual information associated with XYZ gets associated with NEURO RESPONSE X, and vice versa: all scientific information associated with NEURO RESPONSE X gets associated with XYZ.

However, even if one can link and merge the A-slots, the phenomenology is tied to the P-slots. This entails, that when the content of a P-slot consists of monadic markers—since one is unable to fill such a P-slot solely on the basis of the corresponding A-slot—abstract knowledge necessarily leaves out the phenomenology. That is, no matter how detailed abstract knowledge can get via merging with the A-slots of scientific concepts, it remains ineffective in simulating the phenomenology if one lacks the basic constituents—monadic markers—of the perceptual template.

This is the *epistemic gap*, and also the reason why we have the intuition that the physical and the phenomenal are distinct.

Acknowledgement

Bibliography

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Chalmers, D. (1996). *The conscious mind : in search of a fundamental theory*. New York: Oxford University Press.

Chalmers, D. (2003). Consciousness and its place in nature. In S. Stich & T. Warfield (Eds.), *Blackwell Guide to the Philosophy of Mind* (pp. 102-142). Oxford: Blackwell.

Fodor, J. (1987). Why there still has to be a language of thought *Psychosemantics* (pp. 135-154). Cambridge, MA: MIT Press.

Fodor, J. (1998). *Concepts*. Cambridge, Mass.: MIT Press.

Fodor, J. (2008). *LOT2 – The Language of Thought Revisited*. Oxford: Clarendon Press.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly, 32*, 127-136.

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy, 83*, 291-295.

Levine, J. (1983). Materialism and Qualia : The Explanatory Gap. *Pacific Philosophical Quarterly, 64*, 354-361.

Loar, B. (1990). Phenomenal States. In J. Tomberlin (Ed.), *Philosophical Perspectives* (pp. 81-108.). Northridge: Ridgeview Publishing Company.

Margolis, E. (1999). How to acquire a concept. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings*. Cambridge, Mass.: MIT Press.

Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*. New York: Freeman.

Nida-Rümelin, M. (1996). What Mary couldn't know. In T. Metzinger (Ed.), *Phenomenal Consciousness*. Schoenigh: Paderborn.

Papineau, D. (2002). *Thinking about consciousness*. Oxford: Clarendon Press.

Pylyshyn, Z. (2002). Mental Imagery: in search of a theory. *Behavioral and Brain Sciences, 25*(2), 157-237.

Pylyshyn, Z. (2003). *Seeing and visualizing: it's not what you think*. Cambridge, Mass.: MIT Press, Bradford Books.

Shoemaker, S. (1981). The Inverted Spectrum. *Journal of Philosophy, 74*(7), 357-381.

Smith, A. D. (2010). Disjunctivism and Illusion. *Philosophy and Phenomenological Research, 80*(2), 384-410.

Stoljar, D. (2005). Physicalism and Phenomenal Concepts. *Mind and Language, 20*, 469-494.