Introduction: Themes in Ned Block's Philosophy of Mind and Consciousness

Adam Pautz and Daniel Stoljar

1   Four Imaginary Cases

It is often thought that one of the key marks of creativity in the philosophy of mind is the ability to produce effective imaginary or possible cases, cases that illustrate or refute a philosophical thesis. With that in mind, consider the following cases:

Case 1: Blockhead. "Imagine the set of sensible strings recorded on tape and deployed by a very simple machine as follows. The interrogator types in sentence A. The machine searches its list of sensible strings, picking out those that begin with A. It then picks one of these A-initial strings at random, and types out its second sentence, call it 'B.' The interrogator types in sentence C. The machine searches its list, isolating the strings that start with A followed by B followed by C. It picks one of these ABC-initial strings and types out its fourth sentence, and so on. So long as the programmers have done their job properly, such a machine will have the capacity to emit a sensible sequence of verbal outputs, whatever the verbal inputs, and hence it is intelligent according to the neo-Turing Test conception of intelligence. But actually, the machine has the intelligence of a toaster."

Case 2: The China-body system. "Suppose we convert the government of China to functionalism, and we convince its officials that it would enormously enhance their international prestige to realize a human mind for an hour. We provide each of the billion people of China … with a specially designed two-way radio that connects them in the appropriate way to other persons and to [an] artificial body … [equipped with a] radio transmitter and receiver connected to the [body's] input and output neurons.… [Finally] we arrange to have letters displayed on a series of satellites placed so that they can be seen from anywhere in China. The system of a billion people communicating with one another plus satellites plays the role of an external 'brain' connected to the artificial body by radio.… [This artificial body connected to the people and China] could be functionally equivalent to you for a short time, say an hour.… [But] there is prima facie doubt about whether it has … what philosophers have called 'qualitative states.' (You

ask: What is it that philosophers have called 'qualitative states'? I answer, only half in jest: As Louis Armstrong said when asked what jazz is, 'If you got to ask, you ain't never gonna know.')"

Case 3: Commander Data. "Let us think of Commander Data as defined as a merely superficial functional isomorph of us. A superficial isomorph of us is isomorphic to us in causal relations among [its] states, inputs, and outputs to the extent that those causal relations are part of commonsense psychology. That is, for every human mental state, input, and output, there is a corresponding state, input, and output of Commander Data; and for every causal relation among our states, inputs, and outputs, there is a corresponding causal relation among Commander Data's mental states, inputs, and outputs. One consequence is that Commander Data will behave just as we do, as far as we can tell from the standpoint of commonsense psychology. I said that Commander Data is a merely superficial isomorph of us. That means that he is not like us in physical realization of the superficial functional states.… We might suppose just to get an example on the table that the physical basis of Commander Data's brain is to be found in etched silicon chips rather than the organic carbon basis of our brains… We have no conception of a ground of rational belief that Commander Data is or is not conscious."

Case 4: Inverted Earth. "[On Inverted Earth] everything has the complementary color of the color on Earth. The sky is yellow, grass is red, fire hydrants are green, and so forth.… [Here on Earth] a team of mad scientists knock you out. While you are out cold, they insert color-inverting lenses in your eyes, and change your body pigments so you don't have a nasty shock when you wake up and look at your feet. They transport you to Inverted Earth, where you are substituted for a counterpart who has occupied a niche on Inverted Earth that corresponds exactly (except for colors of things) with your niche at home. You wake up, and since the inverting lenses cancel out the inverted colors, you notice no difference at all.… [However,] after enough time has passed on Inverted Earth, your embedding in the physical and linguistic environment of Inverted Earth would dominate, and so your intentional contents would shift so as to be the same as those of the natives."

2   Three Empirical Cases

It is often also thought that creativity in the philosophy of cognitive and neural sciences consists in the ability to connect empirical work with larger philosophical questions, in a way that does justice to both. With that in mind, consider the following cases:

Case 5: Normal variation in color vision. "Color vision varies from one normal perceiver to another. There are three kinds of cone in the retina that respond to long, medium and short-wave light. (Light is electromagnetic radiation in the wavelength zone of 400–700 nm.) The designations 'long,' 'medium' and 'short' refer to the peak

sensitivities…. There are a number of [sex-linked] genetic divisions in peak sensitivities in the population that are analogous to differences in blood types (in that they are genetic polymorphisms, discontinuous genetic differences coding for different types of normal perceivers)…. These differences in peak sensitivities don't show up in common activities, but they do reveal themselves in subtle experimental situations. One such experimental paradigm uses the anomaloscope (devised in the 19th Century by Lord Rayleigh), in which subjects are asked to make two halves of a screen match in color, where one half is lit by a mixture of red and green light and the other half is lit by yellow or orange light…. Whereas one subject may see the two sides as the same in color, another subject may see them as different—e.g., one redder than the other. When red and green lights are adjusted to match orange, women tend to see the men's matches as too green or too red. Further, variation in peak sensitivities of cones is just one kind of color vision variation. In addition, the shape of the sensitivity curves vary. These differences are due to differences in macular pigmentation, which vary with 'both age and degree of skin pigmentation' … Hence races that differ in skin pigmentation will differ in macular pigmentation. There is also considerable variation in amount of light absorption by pre-retinal structures. And this factor also varies with age. I emphasize gender, race and age to stifle the reaction that one group should be regarded as normal and the others as defective. (That would be sexism, racism or ageism.) … My point is that the facts about variation that I have presented give us no reason at all to regard any gender, race or age as abnormal in color vision…. Assuming that most men and women, blacks and whites, old and young have veridical color vision, two experiences can have the same representational content but different phenomenal character."

Case 6: The Gabor patches. "The subject is shown [a] fixation point, then two 'Gabor patches' … If [one] Gabor patch is slightly lower in actual contrast, attention [to it] can boost its apparent contrast to the point of apparent equality of contrast between the two patches. If the two Gabor patches are the same in actual contrast, an attended patch looks higher in contrast…. If changing the distribution of attention changes the representational contents (cluster of properties/sensible profiles), the upshot is that at least one of those representational contents is illusory, [but] that claim is wrong."

Case 7: GK. "There is a type of brain injury which causes a syndrome known as 'visuospatial extinction.' If the patient sees a single object on either side, the patient can identify it, but if there are objects on both sides, the patient can identify only the one on the right and claims not to see the one on the left. (With competition from the right, the subject cannot attend to the left.) However, as Geraint Rees has shown in two fMRI [functional magnetic resonance imaging] studies of one patient (known as 'GK'), when GK claims not to see a face on the left, his fusiform face area (on the right—which is fed by the left side of space) lights up almost as much as—and in overlapping areas involving the fusiform face area—when he reports seeing the face. Should we

conclude that the fusiform face area is not the whole of the core neural basis for the experience as of a face? Or that activation of the fusiform face area is the core neural basis for the experience as of a face but that some other aspect of the total neural basis is missing in GK? However, another possibility is that GK genuinely has face experience that he doesn't know about and cannot know about. Wait—is that really a possibility? Does it even make sense to suppose that a subject could have an experience that he doesn't and can't know about? What would make it his experience?"

3   The Mind behind the Cases

The mind behind these cases—the mind who invented cases 1–4 and the mind who spotted the relevance to philosophy of cases 5–7—belongs to, or perhaps just is, Ned Block.[1]

That there is one mind here is striking. Philosophy of mind is a field subject to two quite different masters. On the one hand, the field needs to do justice to the complexity of philosophical concerns about consciousness, sensations, perception, meaning, attention, and knowledge. On the other hand, the field needs to be responsive to ongoing developments in the sciences of psychology, linguistics, neuroscience, and computer science. Most philosophers of mind will tell you that they are keen to engage with both research programs, but the truth is that usually one of them dominates their thinking. This is not true of Ned Block. Perhaps more than any other philosopher of mind, Block synthesizes philosophical and scientific methods with creativity and intensity. While philosophers often concentrate on one or the other, Block is remarkable for moving back and forth across this divide with ease. In fact, the nature of his work makes the divisions here somewhat artificial.

4   Themes in Block's Work

What sorts of philosophical themes does Block think these cases and other cases illustrate? More generally, what are the main themes of his work? It is difficult to provide any pithy summary of a body of work as varied as Block's. But for readers who are new to his work, it would not be too far wrong to view his work as contributing to three major themes.

Theme 1 is intelligence. Inspired by Alan Turing's landmark paper "Computational Machinery and Intelligence," a number of philosophers and scientists have been attracted to a definition of intelligence in terms of what is sometimes called the Turing test. According to this definition, a subject is intelligent just in case the subject is disposed to pass the Turing test, where a Turing test for these purposes means a test wherein a designated judge could not discriminate between the subject and an uncontroversially intelligent subject for a designated duration.

Ned Block's Blockhead example presented in "Psychologism and Behaviorism" (1981) provides a decisive refutation of this definition of intelligence: Blockhead would pass the test but not be intelligent. (The name "Blockhead" is due to Frank Jackson. With the title for this volume we have introduced a second meaning for this term into the philosophical lexicon: Blockhead, *n*. An enthusiast for the work of Ned Block; compare "deadhead," "cheesehead," "gearhead," and similar terms.) Later, Block (under the influence of Stephen White) pointed out that the Blockhead example refuted not simply the Turing test approach but a very large class of theories in philosophy of mind as well: roughly any theory that is committed to supervenient behaviorism, the thesis that if two subjects are identical in their dispositions to behave, they are identical psychologically (see Block 1995c, note 1, 383). In fact, quite a few influential approaches would seem to fall victim to the Blockhead example, including input-output functionalism, Dennett's Intentional Stance Theory of belief and desire, and Davidson's Interpretationist Theory of belief and desire.[2]

If the approach to intelligence inspired by the Turing test fails as a result of Blockhead, what should take its place? Here Block is sympathetic to a different idea present in Turing, the computational theory of mind (developed by Block's onetime colleague at MIT Jerry Fodor). On this view, for a subject to be intelligent is for the subject to instantiate a particular computational property—though which computational property is a matter for empirical investigation. The computation theory provides a solution to the problem of how intelligence and rationality find a place in a physical world. For computational properties are entirely realized in physical properties even if they are not reducible to such properties. This view also accommodates the *multiple realizability* of intelligence. For instance, the China-body robot and Commander Data are intelligent because (unlike Blockhead) they have the right computational properties, even if they lack the biology of the human brain.

The concept of intelligence also raises a number of moral concerns, especially in its use by some scientists to try to argue that some races are genetically inferior to others in IQ. Block has contributed here as well in his influential paper "How Heritability Misleads about Race" (1995a). The key insight is an application of the distinction between heritability and genetic determination. IQ may be highly heritable even if it is not genetically determined, because genes may affect IQ indirectly by affecting traits (such as skin color) that interact with the environment so as to affect performance on IQ tests.

Theme 2 is representation. How do our beliefs and desires—or the states in the head that realize our beliefs and desires—manage to represent certain states of affairs? In his paper "Advertisement for a Semantics for Psychology" (1987), Block defends a kind of long-arm functionalism. The representational contents of our internal physical states are determined by a combination of (1) their relations (usually causal or informational

relations) to the external situations they represent and (2) their functional role in producing outputs and in interacting with each other. Like his computationalist view of intelligence, Block's functionalist theory of representation allows for multiple realizability. Thus, the China-body robot and Commander Data have beliefs and desires and represent the world because they have the right functional properties, even if they lack the biology of the human brain.

Another issue that Block has contributed to here is the debate over whether the representational format of imagery in the brain is pictorial (iconic) or descriptive (discursive). In "Mental Pictures and Cognitive Science" (1983), Block proposed that the representational format of imagery (and perception) in the brain is iconic, and he defended this proposal from several prominent objections. Today many (e.g., Carey 2009; Fodor 2007) consider iconicity to be one of the distinguishing features of perceptual representation.

Theme 3 is consciousness. Perhaps the most central theme in Block's work has been the nature of consciousness. Many philosophers see consciousness as reducible to other things: functional organization, representation, or cognitive access. Block has long opposed these ideas. In a word, he advocates a kind of *separatism* for phenomenal consciousness.

First, Block holds that phenomenal consciousness is separable from *functional organization*. We have seen that Block is a long-arm functionalist about representation and a computationalist about intelligence. Therefore, he holds that the China-body robot is intelligent and has beliefs and desires because it has the right computational and functional properties, even if it lacks the biology of the brain. But his view of phenomenal consciousness is radically different. Here he thinks that we should *not* accept functionalism. On functionalism, all that matters to phenomenology is function. So a form of functionalism about phenomenal consciousness would predict that the China-body robot has all the same experiences as a normal human, as long as it is a functional isomorph of a human being, despite its "brain" being made up of a billion people communicating with one another. Against this, Block says in "Troubles with Functionalism" (first published 1978; revised 1980) that the China-body robot *should be regarded as not* having any phenomenal consciousness at all, even if it is functionally indistinguishable from a human being, because of the extremely odd physical basis of its functional states. In particular, he says that we have the "intuition" that the China-body robot would not have experiences, and "if there is no reason to disregard this intuition … we are justified in rejecting Functionalism" (1980, 283). It may represent the world and have beliefs and desire and intelligence (289), but it should be regarded as not having any *experience*.[3] Likewise, even if Commander Data has beliefs and desires, Block says that it is at least an open question whether this system has *experiences*.

In fact, one of Block's earliest papers, "What Psychological States Are Not" (1972), cowritten with Jerry Fodor, was also devoted to showing that phenomenal consciousness

is separable from functional organization (at a certain level of description). In this paper, Block and Fodor use the traditional inverted spectrum scenario to argue against functionalism about color experience. They say that it is at least "coherent" that two individuals, Jack and Jill, should have "inverted color experiences" (172) in response to external objects and yet be functionally exactly alike—that is, alike in what stimuli their internal brain states are normally caused by in the external world, in what further internal states they cause, and in what behavioral outputs they cause. If this is a genuine possibility, then there is more to phenomenology than function.[4] On Block's account of this case, Jack and Jill's experiences have inverted *color qualia* ("mental paint"). For Block, color qualia are "properties of experience," and they cannot be defined in functional terms (2007b, 74).[5]

Second, Block holds that phenomenal consciousness is separable from representation. In the 1990s, a radically new approach to phenomenal consciousness came to the fore: externalist representationism (Harman 1990; Dretske 1995; Tye 1995; Lycan 1996). On this view, the qualities that we experience ("qualia") are properties of items in space, just as they seem to be, rather than properties of our experiences. For instance, color qualities are just physical properties of surfaces and volumes involving the reflection of light. To have a color experience of a certain character is to be in an internal state that normally tracks, and thereby represents, the occurrence of a certain physical property involving the reflection of light.[6] Against this view, in "Inverted Earth" (1990), Block argues that phenomenal consciousness is separable from externally determined representational content, which is fixed by causal-informational relations to the environment. This is the lesson he draws from the Inverted Earth thought experiment—the flip side of the inverted experiences thought experiment. Suppose you are transported to Inverted Earth and you are equipped with inverting lenses. Even though the sky is yellow, thanks to the inverting lenses, you have a bluish experience of it, just as on Earth. After a time, according to Block, your bluish experience of the sky comes to indicate or represent yellow light in the external world, whereas on Earth it represented blue light. A simple (nonteleological) form of externalist representationism would predict that you should at this point come to have a yellowish experience of the sky, rather than a bluish one, even if your internal, physical response to the sky has remained exactly the same throughout; the change in the character of your experience would be due to the change in the normal external cause of your experience. That is, externalist representationism leads to *phenomenal externalism*. Against this, Block assumes the traditional position of *phenomenal internalism*: since your internal physical response to the sky stays the same, the "bluish color quale" of your experience must stay the same too, even if its normal cause in the external world has changed (1990, 68). He writes, "We can assume the supervenience of qualia on the brain, building brain-identity into the story" (1994, 518).[7]

The Inverted Earth case is imaginary. However, more recently, in "Sexism, Racism, Ageism, and the Nature of Consciousness" (1999) and "Attention and Mental Paint"

(2010), Block has used the empirical cases of normal variation in color vision and the Gabor patches to argue that phenomenal consciousness is separable from representation. They are new examples of the ancient problem of variation in appearances among normal observers.

For instance, suppose that, because of innate differences in their wiring, a round color chip looks pure blue to Jack and greenish blue to Jill. Block argues that, in this case, the best account of the difference in the character of Jack's and Jill's color experiences is that it consists, not in a difference in how they represent the chip to be, but in a difference in the "phenomenal ways" (or "nondescriptive modes of presentation") in which they perceive the chip (Block 2007b, 89). What kinds of entities are these phenomenal ways, and where in the world do they reside? Block says that they are *color qualia* (2007b, 85). Recall that, for Block, qualia are "*features of experience*" (74; our italics). Now, on Block's biological theory of experiences (discussed below), experiences are internal states of the brain. Putting all this together, Block's view is that the two distinct phenomenal ways the round chip appears to Jack and Jill are, in fact, features of their own internal brain states. His account of this empirical case is similar to his account of the imaginary inverted spectrum case discussed above (see also chapters 24 and 28).[8]

However, while Block opposes the strong representationist thesis that *all* differences in the character of experience (e.g., differences in *color* phenomenology) are differences in the representational content of experience, it should be mentioned that he is open to the thesis that an experience might essentially have a certain representational content in all possible cases. In particular, he is open to the thesis that some types of visual experiences necessarily have a certain *spatial* content in all possible cases (2007a, 538; 1995b, 278; chapters 8 and 24). For instance, he writes, "There is an element of truth in shape primitivism—the view that circlish experiences necessarily present circleness" (chapter 8). He calls his view "weak representationism" (chapter 16).

To illustrate, consider the brain-in-a-vat case (Block 2012). As noted, Block assumes a kind of phenomenal internalism. So he thinks that a lifelong, accidentally created brain in a vat (BIV) could have, for instance, the very same experience you have when you look at a tomato on a table, provided that it has the right brain state (see Block and O'Regan 2012). Now, of course, the BIV's experience is not round (on Block's view, it is just a brain state, which is not round). And there exists no round nonphysical sense-datum, and no round "visual field region" (Peacocke 2008), within the brain or anywhere in its vicinity. There is just the physical brain. Nevertheless, if the BIV really does have the same tomato-like experience as you, there is a sense in which it vividly *seems* to the BIV that there exists a round thing *right there*; the BIV has an experience as of a round thing. It is open to the weak representationist to say that the BIV having a vivid impression that a round item is present consists in nothing but the BIV's experience (brain state) *representing* the presence of a round item, even if no such item is present; as Block says, the BIV is "tricked" (Block and O'Regan 2012, 105).[9] A consequence of this view would be that

experience involves a primitive form of internally determined phenomenal representation that cannot be defined in functional-externalist terms, since the BIV's brain state is not normally caused by the presence of round things and is not apt to cause round-appropriate behavior. On this kind of weak representationism, then, while phenomenal consciousness is separable from representation to some degree (e.g., differences in color phenomenology are not essentially representational differences), phenomenal representation may not be completely separable from representation (some experiences may have built-in, internally determined spatial representational content).

Third, Block thinks that phenomenal consciousness is to some degree separable from cognitive access. In "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience" (2007a) and elsewhere, he uses a kind of iconic memory to support the view that perceptual consciousness is richer (i.e., higher in informational capacity) than cognitive access. That is, when we observe a complex scene, we are conscious of more than we can report or think about. In fact, he thinks that neuroscience may eventually support a theory on which the neural machinery that grounds conscious experience is completely separable from the neural machinery that grounds cognitive access. This would imply the intriguing possibility of totally inaccessible experiences. For instance, on this view, GK (described in case 7 above) genuinely has face experience that he does not know about and cannot know about (Block 2007a, 498). Pretheoretically, this might seem impossible; but science often shows our intuitions to be faulty.[10]

In sum, Block holds that phenomenal consciousness is *not* reducible to functional role, representational content, or cognitive access. But then, what does he think consciousness *is*? What exactly is his positive view? Does he take the next step and accept the dualist view that phenomenal consciousness is separable from everything physical?

No—Block is a staunch physicalist. In particular, he tentatively favors a *biological theory* of phenomenal consciousness. As he puts it in chapter 34, he thinks that there is "probably" something about our biological makeup—perhaps its "electrochemical character"—that is essential to having any conscious phenomenology (see also Block 2009). In fact, experiences are identical with or grounded in internal biological states of the brain. On this view, experiences are separable from superficial functional organization, from externally determined representational content, and from cognitive access, because they are constituted by internal biological states of the brain, and those biological states are separable from all those things. For instance, Jack and Jill could have "inverted internal brain states" and hence have inverted color experiences, even if those brain states have the function of indicating the same reflectance properties in the world and result in the same behavior. This is why spectrum inversion is possible among such functional duplicates, contrary to functionalism and externalist representationism. Or again, a system (like the China-body robot or Commander Data) could lack the relevant brain states, and hence lack experiences altogether, but be functionally identical to a normal human being.