If Ned Block were a rockstar he would be Mick Jagger: sartorial, iconic, ever youthful, and still producing hit records after half a century. Fittingly, then, Pautz and Stoljar’s celebratory collection has the feel of a tribute concert. The billing includes eighteen papers from colleagues, close friends and former students (Geoff Lee and Janet Levin as well as both Editors). And after each act, Block takes the stage to offer an invariably illuminating response. Most contributors explore ‘big themes’ from Block’s back-catalogue. Consciousness is unsurprisingly top of the bill. Geoff Lee, Michael Tye, Brian McLaughlin, and Hakwan Lau and Richard Brown all explore its basis and (in Lee’s case) significance (or lack thereof). And Joseph Levine, Daniel Stoljar, and Nico Silins and Susanna Siegel all examine aspects of Block’s long-standing contention that phenomenal consciousness can occur in the absence of access consciousness and attention. Bill Brewer, Janet Levin, Sebastian Watzl, and William Lycan address Block’s various arguments against naïve realism and strong representationism (the view that phenomenal character can be reduced to representational content), especially those based on Marisa Carrasco’s influential studies of how attention alters appearance which she helpfully reviews in her contribution. Sydney Shoemaker and Frank Jackson discuss physicalism; and Adam Pautz and David Chalmers raise various intriguing issues about spatial experience, riffing on a host of Blockean motifs including left-right reversals, twin earths, inverted earths, illusions and hallucinations. Even those few contributors who do not directly engage with Block’s work, seek—as Tyler Burge puts it—to honour Block as one musician would another, by ‘trying to play music well and in his spirit’ (p. 41; all page references are to papers in the volume unless otherwise stated).
What then is Block’s spirit? To answer this question, there is no better place to start than Block’s graduate advisor and ‘main intellectual influence’ (p. 451), Hilary Putnam. In his paper, Putnam acknowledges the impact on his thinking about perception of Block’s essays ‘Wittgenstein and Qualia’ (2007a) and ‘Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience’ (2007b), comparing it to the impact which Quine’s ‘On What There Is’ (1948) and ‘Two Dogmas of Empiricism’ (1951) had on his thinking about the philosophy of mathematics. In reading their exchange, it is hard not to be moved, both by Putnam’s open and keen-mindedness late into his eighties, and by Block’s evident emotion in replying to him fifty-five years after they first met, and shortly after Putnam’s death.

Block singles out two Putnamian influences. First, his teaching him ‘the utility of science fiction examples in philosophy’ (p. 451). Second, his emphasis on science. Both are evident throughout Block’s corpus and inspire many of the contributions to Blockheads! They serve as natural themes for this review. I begin with science fiction before turning (mirroring, as I’ll suggest, Block’s own development) to science.

1. **Block the philosopher of science fiction**

Like many, my first encounter with Block was through his seminal ‘Troubles with Functionalism’ (1978)—a paper packed with vivid science fiction examples: from a small country whose economy is rigged to functionally mimic a mind, to a futuristic world where our brains continue to control our bodies by radio despite having been removed for cortical cleaning. Two such cases are now canon.
Homunculus Head: a functional isomorph realized by a billion human beings receiving information from one another and from an artificial body’s sensory and motor neurons via two-way radios (and informed by satellite displays of their system’s current state). (1978, p. 279)

Elementary-Particle-People Infestee: after intelligent aliens much smaller than elementary particles build spacecraft to mimic the characteristics of our elements, you become largely composed of such particle-people ‘elements’ whilst exploring their region of the universe. (ibid., pp. 291-2)

To these, Block (2002) adds a third much-discussed case:

Commander Data: a functional isomorph whose functional organization is realized by a silicon-chip brain. (2002, pp. 401-4)

For those unfamiliar with this terminology: A functional isomorph is a system that is isomorphic to us in respect of a certain set of causal relations amongst mental states, inputs and outputs. An isomorph is superficial if the set includes only those relations specified by folk psychology (and whatever these entail). A psychofunctional isomorph is one in which the set includes all those relations specified by a complete psychological theory. Commander Data is initially conceived as a superficial isomorph. If he lacks consciousness, this contradicts superficialism, the view that consciousness is identical with a particular kind of folk functional
organization. *Homunculus Head* can be construed either as a superficial or a psychofunctional isomorph. Considered as a psychofunctional isomorph, its putative lack of consciousness contradicts psychofunctionalism, the view on which mentality in general and consciousness in particular corresponds to a particular kind of psychofunctional organization. *Elementary-Particle-People Infestee* is intended as a *neurobiological* isomorph, differing from us only in microphysics.

Many of the contributors to *Blockheads!* pursue arguments significantly based on such science fiction thought experiments: Jackson, Lee, McLaughlin, Shoemaker, Tye, and in a rather different setting both Chalmers and Pautz. I’ll focus here on Tye’s ‘Homunculi Heads and Silicon Chips’ and McLaughlin’s ‘Could an Android Be Sentient?’.

Tye contends that consciousness is missing in *Homunculus Head* but present in *Elementary-Particle-People Infestee* as well as in a human whose neurons have been replaced one-by-one by functionally-equivalent silicon chips. By inference to the best explanation, he concludes that, contra *superficialism*, history (and not mere function) matters to phenomenology. Tye is unabashedly confident in his intuitions. *Homunculus Head* is not conscious, since essentially the same as a large corporation. (In his view, corporations, like Apple and Google have mentality quite literally, but phenomenology only metaphorically.) This armchair intuition is buttressed by three considerations. First, that ‘surely’ (p. 555) there would be no issue flaying *Homunculus Head*’s artificial body whilst the system was up and running. Second, that it would ‘hardly be tragic’ (ibid.) if the body fell into a coma. Finally, that ‘patently’ (ibid.) there would be nothing wrong in dismantling the system after an hour. Setting distaste at the cavalier violence aside, how can Tye be so sure of such intuitions? His use of ‘surely’ is a good guide to thin ice, a warning bell (as Dennett likes to note) that the author lacks an argument and hopes to achieve agreement by rhetorical force alone. Block makes a similar point regarding Tye’s discussion of his silicon-chip case: ‘At every crucial
point, Tye appeals to “plausibility.” I count ten occurrences of variants of the word, not counting other words with the same use … or considerations of what it is “reasonable to say.”’ (p. 577)

One familiar objection to science fiction cases is that they all too often leave critical details unspecified. In this connection, it is striking that Tye never makes entirely clear what kind of isomorph he takes his Homunculus Head to be. His comparison with corporations suggests a merely superficial isomorphism, and Block (p. 572) appears to read him this way. But this raises the question why a psychofunctional diagnosis of the alleged pattern of phenomenology would not be preferable to Tye’s historical account. After all, both the silicon-chip replacee and Elementary-Particle-People Infestee share our psychofunctional organization. On the other hand, Tye’s later description of the one-to-one correspondence between humans in the homunculus system and neurons, and the system’s functional replication of the human visual system (p. 563) suggests a deeper psychofunctional isomorphism. This makes more sense of his argument. Yet treating Homunculus Head this way obviates the force of the comparison with corporations, and induces further qualms concerning Tye’s confidence that it lacks phenomenology. In consequence, Tye’s conclusion that history matters to phenomenology rests on rather shaky foundations.

McLaughlin considers whether we could have reason to believe or disbelieve that Commander Data was conscious. His argument is roughly this: we could only have a positive reason to generalize from our own case to Data’s if superficialism were true. But superficialism is false; thus, his radical difference in physiological make-up provides a ‘basic reason … for denying that Data is conscious’ (p. 346). McLaughlin gives two reasons to reject superficialism. The first is a direct appeal to Block’s Homunculus Head (treated as a superficial isomorph) which he takes to lack phenomenal consciousness (p. 354). It is unclear why. McLaughlin rightly draws attention to our ‘System 1’ ‘hair trigger’ mechanisms (p. 340) which
can easily yield false positive diagnoses of mentality when confronted with biological motion or facial features. But we are also notoriously prone to ‘out-grouping’ or ‘dehumanizing’ when focusing on differences (e.g., Harris and Fiske 2006, 2009, 2011). It is not, then, implausible to suppose that the sheer ‘otherness’ of *Homunculus Head* biases us against attributing it consciousness.

McLaughlin’s second, more developed objection to thinking of *Homunculus Head* as conscious concerns absent role cases involving phenomenal states in the alleged absence of corresponding folk functional states. Here McLaughlin offers real-world cases which are both important and interesting. However, they are also inconclusive. Some cases fairly clearly do involve phenomenology, namely anaesthetic awareness and total locked-in syndrome. Yet these are not straight-forwardly absent role cases, since (as McLaughlin acknowledges) causal roles involving other mental states are present. Pain which – due to paralysis – one can do nothing to prevent can understandably have wide-ranging and traumatic mental effects. To this, McLaughlin replies: ‘Cases abound in which the causal role vis-à-vis other mental states is truly sparse’ (p. 357; cf. Lee p. 224). But in these cases, it is much less clear what phenomenology is in fact present. I doubt, for instance, that our intuitions about the conscious mental life of a neonate whose ‘autonomic system is entirely shut down’ (p. 357) are robust.

As Block notes (p. 375f.), there is also a standard reply to absent role cases which appeals to Shoemaker’s notion of ‘paradigmatic embodiment’ (Shoemaker 1976, p. 115; Block 1978, p. 298f.). Start with a functional characterisation of a paradigmatically embodied individual, P (e.g., an ordinary adult human). Then say that another individual, I (e.g., a brain in a vat), is in the same mental condition as P just insofar as the putative realizer of I’s mental life (e.g., brain) could become incorporated into an individual of the same kind as P without altering its internal structure or state relations. McLaughlin’s case against superficialism is thus indecisive, and with it his argument for denying Data sentience.
Both Tye and McLaughlin’s approaches raise a general worry about methodology. Science fiction examples have the large advantage of not requiring expensive equipment or laborious data collection. On the other hand, science fiction is a form of story-telling, and as Wiggins observes, ‘It is perfectly notorious that not every story corresponds to a possible world.’ (2001, p. 66) One’s easy enjoyment of other Star Trek conceits such as teletransportation, warp drives, mind melds, or time travel in no way support these being genuine possibilities. This worry also prompts the question of where science fiction fits alongside Block’s other deep, Putnamian commitment to serious science?

One key point is that Block is vastly more cautious about the force of thought experiments than many of his commentators. Block acknowledges that Homunculus Head is ‘not an overwhelmingly powerful argument’ (1978, p. 296). It carries weight only when backed-up by a positive argument against folk functionalism and in the absence of any ‘minimally decent argument’ (ibid.) for folk functionalism. Likewise, he holds that whilst ‘there is a prima facie doubt whether there is anything which it is like to be the homunculi-headed system … prima facie doubt is only prima facie’ and that ‘appeals to intuition of this sort are notoriously fallible’ (1978, p. 281) and ‘hardly to be considered bedrock’ since they ‘soon disappear’ when confronted with a ‘well-supported theory’ (ibid., p. 293). In fact, Block appears ambivalent about the import of such cases. In replying to McLaughlin, Block writes: ‘I reject the claim that we have a conception of how to find out that homunculi heads are phenomenally conscious.’ (p. 377; see also his 2002, p. 413 and the Editors’ introduction, p. 14) On the other hand, in replying to Tye, Block states that Homunculus Head ‘has no phenomenology … because it lacks the biological mechanisms that underlie phenomenology’ (p. 572). Critically, however, even if Block does deny Data and Homunculus Head consciousness, his commitment rests on rational, theoretical grounds not mere appeals to intuition.
What grounds? The first is that, ‘mentality is in the domain of psychology and/or physiology’ (1978, p. 301); the second is that *Homunculus Head* ‘need not have either psychological (information-processing) or physiological mechanisms anything like ours’ (ibid.). This argument immediately raises the question of what to say about a duplicate which was psychofunctionally isomorphic to a human, that is: a duplicate of which completed human psychological theory would be true. Here, Block argues that there remains a ‘prima facie doubt that it has qualitative mental states’ (ibid.). Again, recognizing that this intuition requires a rational basis, Block offers ‘an argument that qualia are not in the domain of psychology at all’ (ibid.). This argument rests on the claim that nothing in currently conceivable psychology explains qualia (ibid., p. 307).

Block does not think that physiological duplicates raise the same prima facie doubts. This is what *Elementary-Particle-People Infestee* is supposed to show: Our subjectivity would survive such infestation because it would not affect our physiology. But suppose a theorist had a different intuition about *Elementary-Particle-People Infestee*. Troublingly, this theorist seems able to offer an exact analogue of the argument Block gives in support of his intuition concerning the psychofunctional duplicate. Nothing in currently conceivable physiology explains qualia, she might say. Why, she might press, should it be like anything for a subject whose brain manifests neural firing patterns of such-and-such kinds? She might indeed cite Block: ‘We do not see how to explain a state of consciousness in terms of its neurological basis.’ (2002, p. 391) To this critic, Block might reply that we know *we are brain-headed systems*, and that *we have qualia*’ (1978, p. 293) But what makes the biological level the appropriate level at which to make such a gambit? Why not: *We are creatures built from elementary particles, and we have qualia*? Or jumping up a level: *We are information processing systems, and we have qualia*? It remains puzzling then what precisely the grounds are for preferring a biological theory. Indeed, ceteris paribus, should we not prefer the boldest,
most explanatorily general, high-level theory consistent with the data? If so, psychofunctionalism would seem the better bet.

This last thought leads to a second concern, however. Block assumes that if qualia are not in the domain of psychology then we must look *downwards* to physiology. But supposing Block is right that qualia are missing from psychology, what if this is because consciousness is more like being *witty or garish or virtuous*—a property which whilst perfectly real is not a scientific kind at all? After all, consider someone who doubted whether one could explain what made a particular aside so witty by appeal to facts about information processing. Would this sceptic be any more likely to be impressed with an explanation in terms of neurotransmitters? Here compare Elizabeth Irvine’s (2012) claim that science should abandon the concept of consciousness, a claim which as she notes “may not of course translate into worries about the concept in other settings” (2017, p. 96), as well as the discussion of manifest and psychological kinds in Ian Phillips (2018, pp. 476-8).

A second important point to note regarding Block’s use of science fiction cases is that such appeals are always to some degree empirical. The claim that our qualia are brain generated is empirical, as is the claim that we are evolved not designed (Block 1978, pp. 293-4). Similarly, in advocating Shoemaker’s paradigmatic embodiment strategy, Block underscores that determining whether a system can be incorporated unaltered into a paradigmatically embodied individual is a fundamentally empirical matter (1978, p. 299).

Putting these points together, it is tempting to see Block’s science fiction thought experiments as invitations to experimental investigation as opposed to a method unto themselves. Certainly, Block’s own work has turned increasingly from such cases towards the cognitive neuroscience of consciousness. (Consider also Block’s reply to Stoljar where he comments on his shift in thinking about access-consciousness from ‘characterizing a strand of
ordinary thinking about consciousness’ to ‘characterizing an information-processing image of phenomenal consciousness’ (p. 538). Many contributors to *Blockheads!* follow suit. In particular, the papers by Burge, Lau and Brown, Pautz, Silins and Siegel, and Watzl are all heavily reliant on empirical studies. Let us then consider Block the philosopher of science.

2. **Block the philosopher of science**

Block’s love and knowledge of a broad sweep of scientific work courses through his eighteen replies. Block cites evidence of colour categorization in honey bees against Shoemaker’s phenomenal character-based approach to colours; evidence of susceptibility to geometric illusions in four-day-old chicks as problematic for Brewer’s approach to illusions; evidence of radical cross-cultural variation in colour preferences as a challenge to the analytic functionalist; and studies of anosognosia for hemiplegia (the systematic denial of paralysis on one side of the body) as ‘useful corrective to the idea [advocated by Chalmers and Tye] that there cannot be a massive disconnect between phenomenology and belief about phenomenology’ (p. 576). Testifying to her major influence on his work in recent years, the shortest of Block’s replies is to the neuroscientist Marisa Carrasco: a gracious, ‘thank you’ (p. 107).

In contrast to its fictional counterpart, real world science has the major disadvantage of being fraught with technical, methodological and interpretive difficulties—all the more so when it remains in its adolescence, as with psychology. Despite this, Block’s work consistently demonstrates the philosophical value of careful and informed engagement with relevant empirical work.
A nice illustration of this is the illuminating exchange between Lau and Brown, and Block. Lau and Brown are proponents of a higher-order thought (HOT) theory of consciousness, the proposal that a first-order state is conscious in virtue of being the object of a higher-order representation that one is in the state (Rosenthal 2005). A question which quickly arises for such views is what to say about cases of empty higher-order representation. If one claims that the higher-order representation determines one’s conscious experience even in the absence of a first-order representation, then it seems that it is the higher-order state which is doing all the work. Indeed, the initial idea that first-order states become conscious appears to lapse—in what sense are they conscious when they would seem neither necessary nor sufficient for a subject’s conscious experiential condition (Block 2011; Byrne 1997)? Against this, Lau and Brown raise three cases which they contend involve actual examples of phenomenal consciousness constituted by empty higher-order representations. These cases, urge Lau and Brown, show that it is the first-order theorist who is in trouble.

By looking to empirical cases to make philosophical progress, Lau and Brown are here explicitly ‘employing a strategy for which Block himself is well-known’ (p. 172). This is laudable—those familiar with these debates may sympathize with Block’s remark that he has found earlier discussions ‘rather unproductive’ (p. 199). The three cases which Lau and Brown consider are also fascinating in their own rights.

(i) Rare Charles Bonnet Syndrome patients suffer from damage to occipital regions, including primary visual cortex (V1), but nonetheless undergo visual hallucinations. Lau and Brown argue that since first-order representations ‘critically depend’ (p. 173f.; citing Lamme 2006, Block 2005, 2007) on intact primary visual cortex, these
hallucinations must be the result of higher-order mechanisms in the absence of first-order representation.

(ii) **Inattentional inflation.** In Rahnev et al. (2011), participants were presented with grating patterns in attended and unattended regions of their visual fields. The luminance of these patterns was adjusted such that participants exhibited equal objective sensitivity to the gratings. Nonetheless, participants were more liberal in reporting the unattended patterns. Lau and Brown suggest that this provides a case where first-order representations, as indexed by objective sensitivity, are ‘not strong enough to account for the conscious experience’, as indexed by report. Higher-order representations are required to bridge the gap or ‘degree of emptiness’ they discern (p. 174).

(iii) **Peripheral Vision.** Lau and Brown claim that the subjective richness of peripheral vision lacks sufficient support in early processing, and so again must be understood in terms of subjective inflation driven by higher-order representations.

As Block brings out, however, the devil is in the detail. The cases of Rare Charles Bonnet Syndrome which Lau and Brown point to do not involve complete destruction of V1. The damage is partial, sometimes unilateral. Moreover, the first-order theorist can simply deny that first-order representations ‘critically depend’ (p. 173f.) on intact primary visual cortex. Extrastriate regions may partake in local recurrent loops which Lamme and Block favour as the basis of (first-order) phenomenal consciousness.

As regards **Inattentional Inflation**, Lau and Brown assume that criterion effects (i.e., effects on one’s tendency to favour a particular response independent of one’s perceptual sensitivity) imply a change in subjective perception. Yet such effects may equally well be—and indeed traditionally would be considered to be—effects on judgement not perception (see,
e.g., Green and Swets 1966, pp. 118-9, and discussion in Phillips 2020). Similarly, citing Green and Swets, Lau and Brown claim that standard detection theoretic models hold that ‘subjective perception happens when [a] signal crosses a threshold or criterion’ (p. 180). But this is only true if ‘subjective perception’ means reported perception which is not what is at issue. Finally, Rahnev et al.’s explanation of inattentional inflation in terms of the greater variance of peripheral signals is entirely consistent with an account on which peripheral vision induces a more liberal decision bias. Block offers a nice analogy to illustrate:

Suppose there are two political demonstrations, one on your left, the other on your right. You want to judge which is louder. You listen for really loud noises. Unknown to you, the average noise level is the same, but the one on the right is more unruly—that is, there is more variability. So the demonstration on the right is more likely to cross the threshold for really loud noises, and you are more likely to judge it to be louder even though the average noise levels are the same. (p. 208)

Turning to peripheral vision, Block provides numerous reasons to be sceptical of the claim that processing in early areas is insufficient to support phenomenal richness. A third of cells in the peripheral retina are cones (Tyler 2015), and spatiotemporal integration of information from such cells, perhaps together with information from less peripheral cells, may well produce a rich colour representations (Pelli and Tillman 2008; Huang, Treisman and Pashler 2007). Furthermore, learned associations within the visual system (e.g., between shape or texture and colour) may supplement relatively poor visual input (Deroy 2013). Finally, even if ‘filling in’ occurs in a top-down fashion such causal influences do not imply higher-order regions are constitutive of phenomenal consciousness.
Block here splendidly illustrates the fraught path from empirical data to theoretical conclusions. Of course, that same critical eye can also be turned on Block’s own empirically-based arguments. In closing, let me offer one case in point. In his reply to Janet Levin, Block adds to her concerns about strong representationism with an additional argument. This argument is premised on the claim that we can find two structurally identical representations one with phenomenal character, one without. The case he proposes is Milner and Goodale’s famous form agnostic patient, DF. DF is consciously able to see colour but not orientation. Testifying to this, DF is close to chance in judging or ‘matching’ the orientation of a randomly oriented slot when the slot is presented at a distance. However, when asked to post a card through the same slot, she performs almost as well as a neurotypical perceiver (Milner et al. 1991). Alongside other findings, this remarkable apparent dissociation of perceptual judgment and action has convinced many theorists that a whole stream of perceptual processing subserving online motor control proceeds wholly outside awareness. This larger contention aside, Block claims that DF supplies a clear case of structurally identical representations (viz., colour and orientation) where only one has phenomenal character.

There is, however, an alternative possibility: DF may be successful in posting by using an orientation-free ‘obstacle avoidance’ strategy, wherein she attempts to minimise the relative distances between card and slot edges (Hesse, Franz and Schenk 2011). Block dismisses this criticism as ‘weak’ (p. 274), pressing two issues. First, that Hesse et al.’s finding that (some) neurotypical subjects exploited an obstacle avoidance strategy in a modified posting task shows nothing about DF. Second, that DF ‘moved her hand forward unhesitatingly, and … smoothly’ (Milner and Goodale 2005, p. 20).

Do these points undermine Hesse et al.’s objection? I do not think so. First, even though only suggestive, Hesse et al.’s findings show that we should take seriously an alternative hypothesis regarding DF’s performance on which she simply has no representation of
orientation. Second, DF’s early and smooth motion does not preclude an obstacle avoidance strategy. The principal difference in strategies is simply in the visual information used. Indeed, Hesse et al.’s data suggest that the neurotypical subjects who adopted an obstacle avoidance strategy exhibited both early adjustment—‘the final orientation of the card is already determined before half of the movement distance is covered’ (2011, p. 160)—and smooth action: Ss were instructed ‘to put the card in a quick and uninterrupted movement through the slot’ (p. 156) and ‘as smoothly as possible’ (p. 158) and made very few errors in either closed or open-loop conditions.

Of course, being science not fiction, these are treacherous waters, and there is undoubtedly much more to be studied and said. The same is true of other putative cases of unconscious perception which Block cites in replying to Lycan (pp. 327-8) and Shoemaker (p. 485)—see Block and Phillips (2016) and Peters et al. (2017) for discussion. But wherever the science eventually takes us we should all learn from Block’s example, and as Burge puts it, his virtuoso capacity for ‘interpreting and clarifying good science’ (p. 41).

IAN PHILLIPS

Johns Hopkins University

ianbphillips@jhu.edu
References


—— 2007a, ‘Wittgenstein and Qualia’, in *Philosophical Perspectives* 21

—— 2007b, ‘Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience’, in *Behavioral and Brain Sciences* 30

—— 2011, The higher order approach to consciousness is defunct’, in *Analysis* 71

Byrne, Alex 1997, ‘Some Like it HOT: consciousness and higher-order thoughts’, in *Philosophical Studies* 86


Harris, Lasana T., and Fiske, Susan T. 2006, ‘Dehumanizing the lowest of the low: Neuroimaging responses to extreme outgroups’, in *Psychological Science* 17

Dehumanized Perception: A Psychological Means to Facilitate Atrocities, Torture, and Genocide?’, in Zeitschrift für Psychologie 219


—— 2020, ‘Blindsight is qualitatively degraded conscious vision’ in Psychological Review, online early <http://dx.doi.org/10.1037/rev0000254>


Quine, Willard Van Orman 1948, ‘On What There Is”, in Review of Metaphysics 2


Tyler, Christopher 2015, ‘Peripheral Color Demo’, in i-Perception, 6