A Role for Qualia

William J. Rapaport

**Department of Computer Science and Engineering,
Department of Philosophy, Department of Linguistics,
and Center for Cognitive Science
University at Buffalo, The State University of New York,
Buffalo, NY 14260-2500**

rapaport@buffalo.edu
http://www.cse.buffalo.edu/~rapaport/

September 5, 2022

Abstract

If qualia are mental, and if the mental is functional, then so are qualia. But, arguably, qualia are *not* functional. A resolution of this is offered based on a formal similarity between qualia and numbers. Just as certain sets "play the role of" the number 3 in Peano's axioms, so a certain physical implementation of a color plays the role of, say, red in a (computational) cognitive agent's "cognitive economy".

1 Qualia, the Mental, and the Functional

Qualia are one of the sticking points in trying to develop a theory of consciousness. This includes computational theories of consciousness, whose abstract formulations (as algorithms) are functional, and whose physical implmentations are in computers (e.g., robots). Consider the following inconsistent triad:

1. The mental is functional.

2. Qualia are mental.

3. Qualia are not functional.

Proposition 1 is intended to summarize the idea that our best current theory of mental phenomena is some form of functionalism:

> Functionalism is the doctrine that what makes something a ... mental state ... depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulations, other mental states, and behavior. (Levin 2021; see also Maley and Piccinini 2013; Piccinini 2020, Ch. 4 for surveys of other varieties).

Proposition 2 asserts a standard view that qualia are "mental state[s] with a very distinctive subjective character" (Tye, 2021). Like all mental states, they are private in the sense that only the cognitive agent experiencing them "knows" what they are like—what it is like to experience them: "[M]ental states are private in the sense that only those who possess them can know them directly. ... [N]eural activities are public in the sense that anyone suitably placed can observe them" (Calef, nd). And, as mental phenomena, what is important about them is the role they play "in the cognitive system of which [they are] a part"—the agent's "cognitive economy".

Proposition 3 is one of the fundamental problems concerning the nature of qualia, as embodied in the problems of the inverted spectrum and absent qualia (Block, 1978) and the problem of the explanatory gap (Levine, 1983). (See §4.2, below.)

The first two propositions imply the denial of the third; yet all three seem to be true. Something has to give.

I will argue that the question "What are qualia?" is formally similar to one of the principal questions in the philosophy of mathematics: What are numbers? The problem of the nature of qualia in theories of consciousness is akin to the problem of the nature of numbers. And I will suggest that the same kind of answer can be given to both questions.

But first some background ...

2   "Easy" (Psychological) Problems of Consciousness

According to David Chalmers (1996, p. 26), among the *psychological* problems of consciousness are awakeness, introspection, reportability, self-consciousness, attention, voluntary control, knowledge, and awareness (Chalmers, 1996, pp. 26–28). An alternative list adds "the ability to discriminate, categorize, and react to environmental stimuli; the integration of information by a cognitive system; ...; the ability of a system to access its own internal states; [and] the focus of attention", *inter alia* (Chalmers, 1995). Ned Block's notion of "access" consciousness—"availability for use in reasoning and rationally guiding speech and action" (Block, 1995, p. 227)—seems to overlap significantly enough with Chalmers's notion for

us to identify them, at least for present purposes. These are what Chalmers calls the "easy" problems, i.e., those explainable, in principle, in abstract (e.g., functional or computational) terms or in concrete (e.g., neural) terms.

Gualtiero Piccinini argues that, at least for animals, the abstract and the concrete are not *disjoint* levels of analysis. Rather, there is a single level of analysis (Piccinini, 2020, §1.6) that involves both "mechanistic functionalism *plus* neural computation" (Piccinini, 2020, p. 106, my italics). This is consistent with what I shall say later (§7.1). Nevertheless, I also hold that the "upper" portions of such a single-level analysis are consistent with—can be implemented in—non-biological (non-neural) mediums, including appropriately programmed computers (Rapaport, 1999, 2000, 2005b, 2012, 2018).

There are several models of psychological consciousness: multiple drafts (Dennett, 1992, 2005), global workspace (Baars, 1988; Dehaene and Naccache, 2001; Anderson, 2007; Baars and Franklin, 2009), higher-order theories (Block, 2001; Rosenthal, 1986, 2009; Kriegel, 2009), etc. The main point is that any of the various models of psychological consciousness could be concretely implemented, neurally or computationally[1]—at least, once most of the details have been worked out (not, of course, an easy task, despite Chalmers's characterization). In any case, even if cognition and consciousness are comput*able*, it does not follow that *human* (or other animal) cognition and consciousness are comput*ed* (Shagrir 1997; Rapaport 1998; Piccinini 2020, pp. 146–147; Chirimuuta 2021).

3   "Hard" (Phenomenological) Problems of Consciousness

Psychological theories of consciousness tell us how consciousness might *work*. But they don't tell us how consciousness *feels*, what an experience is, how and why we feel experiences, what purposes experiences serve, what it's like to be conscious:

> I can explain to you what love is until I turn blue in the face. I can take two weeks to explain everything to you . . . . But there is no way I can make you feel it until you feel it. (Rohail Hyatt, quoted in Schmidle 2008: 40, and sounding for all the world like Frank Jackson (1982, 1986) in his black-&-white-Mary days.[2])

In short, they don't explain qualia. Chalmers (1995) has dubbed this the "hard" problem; Levine (1983) calls it the "explanatory gap".[3]

---

[1]For a bibliography of some computational implementations, see https://cse.buffalo.edu/~rapaport/719/csnessrdgs.html.

[2]I.e., pre-Jackson 1995.

[3]See Mehta 2022 for other characterizations.

There is an important reason for this: It seems that they can't, for logical reasons, in particular, the inverted-spectrum problem. *Inversion* of qualia is not strictly needed for the argument. It suffices if your qualia differ from mine (see Block 1999). But I will continue to speak of inversion for the sake of simplicity and continuity with previous discussions.

The logical possibility of inverted spectra is only one aspect of this problem. More generally, the problem is to understand how the mechanisms of any of the psychological theories end with our experiencing qualia. I can understand that I see you because light reflecting off your surface enters my eye and is processed by my retina and optic nerve, and then signals are sent to my visual cortex. But how is it that the neurons that are firing there are visually experienced by me as you? And why aren't they auditorially experienced as, say, middle C? This hard *phenomenological problem* of consciousness seems to require a textitphenomenological theory of consciousness over and above the psychological one.

Dennett demurs: It might be the case that psychological theories of consciousness don't (have to) explain qualia because there is nothing to explain. Qualia are (supposed to be) ineffable: If they can't be described in language, then we don't *have* to explain them, because we *cannot* (see Dennett 1988): There are no qualia; rather, it only *seems as if* there are, in just the same way as there is no fact of the sun revolving around the earth—it only *seems as if* it does (Pylyshyn, 2003, p. 117, crediting Wittgenstein). On this view, if there is anything to explain, it is only why things seem that way.

But many others (e.g., Nagel 1974; Searle 1992; Crick 1994; Block 1995; Chalmers 1996; Strawson 2012; McGinn 2021) insist that it is obvious that we do experience qualia and that it must be explained. Some (e.g., Crick), after agreeing that there are qualia, then go on to propound only a psychological theory, hoping that the qualia problem will get settled as a by-product. Others (e.g., Chalmers and Strawson) try to attack the problem directly, taking experience as primitive or fundamental. Yet others (notably Jackson) used to believe in qualia but have changed their mind.

This is a dividing issue: One either agrees that there are qualia and that they need to be explained, or one doesn't. I doubt that any amount of arguing will convince anyone on either side (other than Jackson) to switch to the other. I am, as Jackson once was, a "qualia freak" (Jackson, 1982, p. 127). And, following Hector-Neri Castañeda, I prefer to do philosophy "in the first person, for the first person" (Rapaport 2005a; Rapaport 2023, Ch. 2). So I will simply take a stand in favor of qualia, and proceed to try to resolve the inconsistent triad.

Could phenomenological consciousness (qualia) be nothing but neuron firings? Crick's "astonishing hypothesis" is that that's all they are. Searle says that they must be biological, just like digestion is biological. I agree that they must be bio-

logical (neural) ... *for us*. But I believe that they can be implemented otherwise for computational cognitive agents (Rapaport, 1985, 1986, 1988, 1999, 2023). That means that there has to be a functional-computational ("psychological") theory of qualia, despite the apparent logical problems. And this is what I shall try to convince you of.

Briefly (and very roughly), I will argue that, with respect to the "easy" problem, there is an analogy between functional-psychological theories of consciousness and mathematical structuralism and that, with respect to the "hard" problem, there is an analogy between qualia and objects that can play the role of numbers in structural theories of mathematics.

4 The Problem with Qualia

A quale is a "raw feel", a "phenomenal experience", "what it's like" to be conscious of something. It's what you experience when you sense a color, hear a sound, or taste, or smell, or touch. It's what Chalmers says needs to be explained.

But standard *functional* theories of mental states and processes seem incapable of explaining qualia. And standard *psychological* theories of consciousness (as opposed to *purely* neurological theories of consciousness) are functional.

4.1 Functional Analyses

Functional theories characterize mental processes in terms of causal and logical relations *between*, at the outset, sensory *inputs* and *internal* mental states/processes, then *among* those internal mental states/processes, and, finally, *between* internal mental states/processes and motor *outputs*. The standard technique for doing this is what is variously called "functional decomposition" (akin to recursive definition), "top-down design" (Mills, 1971), "stepwise refinement" (Wirth, 1971), "homuncularism" (Fodor, 1968; Dennett, 1975; Lycan, 1981), or the delegation of authority within hierarchies (Simon, 1996): One describes a task or process in input-output terms (as a mathematical function), and then (algorithmically) analyzes the "big" task into "smaller" tasks, each of which is more "manageable" or "understandable" and which are such that, taken together, they accomplish the big task. In the simplest case, a task F that takes input I and yields output O might be analyzed or "decomposed" into n smaller tasks $F_1, \ldots, F_n$ such that F_1 takes as input F's input I and yields output O_1, F_2 takes as input its predecessor's output O_1 and yields output O_2, ... , F_n takes as input its predecessor's output and yields F's output O. Subsequently (by recursive stepwise refinement), each F_i can be similarly decomposed. The decomposition stops with processes that, to paraphrase Dennett,

just say 'yes' or 'no' (or are otherwise "minimal" in some sense). If each F_i is computable, then so is the original F. (These are the first two of Marr's (1982) three-layer methodology; see Rapaport 2023, Ch. 16.)

Although this top-down explanation is the usual way of presenting functional decomposition, looking at it from the bottom up is useful, too: Consider a programming language with a set of basic (non-decomposable) instructions and a set of "control structures"—i.e., grammatical rules—for recursively composing them into complex instructions. For a Turing machine, the basic instructions might be "read-value-on-current-square", "move-left-1-square", "move-right-1-square", "print-0-on-current-square", "print-1-on-current-square", and "erase-current-square", and the grammar rules might be sequence ("linear concatenation", or "**begin** $S; S'$ **end**"), selection ("conditional branching", or "**begin if** value-on-current-square $= 0$ then S else S' **end**"), and while-loop ("**begin while** value-on-current-square $= 0$ **do** S")—where S, S' are (basic or complex) instructions. Now suppose that there is a complex instruction S^*, perhaps involving numerous occurrences of basic instructions composed via numerous instances of sequence, selection, and looping, that occurs repeatedly in some larger program. Good structured programming practice would "factor" S^* out so that there is only one occurrence of it, and replace it by a "procedure call" (a "jump" or "go to") to S^*. Even better, if S^* is given a (descriptive) name, then we will have created a new "quasi-basic", non-decomposable instruction whose "implementation details" are hidden from the main body of the program and which can be replaced by any input-output-equivalent instruction without changing the main body of the program.

Here is a simple example: The Karel-the-Robot programming language (Pattis et al., 1995) is a language for a software robot named 'Karel' that can move along a graph-paper-like grid. One of its basic instructions is "turnleft", which causes Karel to rotate $90°$ counterclockwise. The language does not, however, have a basic instruction called "turnright". Instead, "turnright" can be defined as (roughly) "**begin** turnleft; turnleft; turnleft **end**". One could instead have defined "turnright" as a sequence of 7 "turnleft"'s; if so, one would not need to change any program that required Karel to "turnright". One can then write Karel programs that instruct Karel to "turnright" without worrying about how he does it.

The power of this kind of "procedural abstraction" is that one can create new, complex instructions. By *naming* these new, complex instructions, one can conceive of them as new, "basic" instructions. By making the names descriptive of their behavior, one can create programs that are abstract in the sense that they can be implemented in different ways. Thus, whereas functional decomposition analyzes a complex behavior into smaller behaviors, procedural abstraction synthesizes complex behaviors from smaller ones, and considers the functional behavior of those larger procedures independently of their implementation.

4.2 The Classic Conundrum of Qualia

However, it is difficult to characterize qualia functionally because of the possibilities of inverted and absent qualia.

Let us suppose that neither you nor I are red-green color blind and that we both are looking at a traffic light whose top light is illuminated. Let us suppose that we both say that we see a red light. We agree that it is red, that it is properly called 'red', and that it is (roughly) the same color as the stereotypical fire engine. We also agree that it is not the same color as the bottom traffic light, which we both agree is green, is properly called 'green', and is (roughly) the same color as stereotypical grass. From a functional point of view, we agree on colors.

Now, what about our qualia, our "internal", private color-experiences? Are they the same? Or could it be that the quale you have when we both look at the top traffic light is not the same as the quale I have? Could it be, for example, that our "internal spectra" are inverted with respect to each other, so that the quale that *you* experience when the light is what we both call 'red' is what I would call 'green' and my green-quale is what you would call 'red' (if, *per impossibile*, we could access each other's qualia)?[4]

The classic conundrum of qualia is that, from the functional point of view, this would seem to be perfectly logically possible, yet experimentally indetectable. Moreover, it seems equally logically possible and experimentally indetectable that you might not have *any* quale when looking at the top traffic light, yet you would, like a philosophical zombie, behave as if you did. Whether any of this is *physically* possible would seem to be irrelevant to the *logical* point. And the further problem is that qualia are classically taken to be stereotypical of a mental phenomenon. So, either functionalism is wrong (or at least incomplete) as a theory of mental phenomena or else—*pace* the stereotype—qualia are not mental phenomena (or don't exist at all). In fact, from the functional point of view, they don't seem to be needed: It's only the external *behavior* (the *talk* of red and green, not the *experience* of red and green) that needs to be explained. Yet qualiaphiles demand such an explanation.

Wittgenstein's take on this is that qualia are "beetles in boxes":

> Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. —Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. —But suppose the word "beetle" had a use in these people's language? —If so it

---

[4]Jacquette (1996) argues that it's impossible.

would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a something: for the box might even be empty. —No, one can "divide through" by the thing in the box; it cancels out, whatever it is. (Wittgenstein, 1958, I, §293)

That would certainly be Dennett's position. But I insist that I, at least, do have a beetle in my box—even if mine is red where yours is green, even if Dennett's is absent, even if I may be misled by mine or misremember it. So: I do experience a quale of red when looking at the top traffic light. What is it that I experience? What is experience itself? If I try to analyze it in terms of other aspects of my psychological "mental economy", it loses its "raw feel" nature. Yet, if I take it as a "primitive" that cannot be explained, how can I understand it? I want a theory of qualia that is consistent with a functional theory of consciousness.

5 Towards a Functional Theory

Why (and how) do we experience anything rather than nothing? This "philosophical zombie" question (Kirk, 1974; Chalmers, 1996) is the *really* hard problem. And why are our qualia as they are and not like something else? Why, that is, is my red quale the way it is and not like my green quale? For that matter, why isn't it like the sound of a bell? Answers to these may depend on the answer to the first question.

So let's begin with that first question: Why do we experience anything rather than nothing? Why are we not zombies? Consider plants. Some are sensitive to light, some to airborne chemicals (what we would call odors) (Angier, 1992; Fountain, 2006). Why don't we say (except metaphorically) that such plants "see" or "smell"? Why *do* we say that *we* see and smell? Or *do* we see and smell? Perhaps we, too, are merely light- and chemical-sensitive (and perhaps that's all that seeing and smelling are).[5] Does anything *further* happen in our brain that does not happen to a plant? If so, wouldn't whatever that is be a good candidate for a quale? Or are qualia just our sensitivity to the light and chemicals? Or is there any sensitivity (or sensation) at all?

Consider the visual quale of seeing a red traffic light and the olfactory quale of an odor. And compare these with the *lack* of an olfactory quale: We are *qualitatively in*sensitive to many odors. (Better: We are qualitatively smell-insensitive to many airborne chemicals.) We are certainly less sensitive than a dog. Yet possibly these odors do influence our behavior. Indeed, there is a phenomenon called "blind smell" that is analogous to blindsight (Sobel et al., 1999; Chen and Haviland-Jones,

---

[5]Perhaps like certain blind worms that respond to blue light (Greenwood, 2021).

2000). If this is indeed possible, then it would seem that we are at least *partial* zombies. How can this be? Why is there such a difference between sensory input that produces qualia and sensory input that doesn't? Does the (visual or olfactory) quale *do* anything? Does it serve any purpose?

Jackson (1982) famously argued that they are epiphenomenal, but I have my doubts. Compare your visual quale of a red light with your *absence* of a visual quale of infrared light. Suppose infrared light influenced your behavior but you were not *subjectively* aware of the infrared light. Now, you *could* be *objectively* aware of it: You might have access to a device that is sensitive to infrared light, so that you would know when you are being exposed to it even though you could not sense it. Or, you might have access to a device (perhaps an "autocerebroscope" or some kind of fMRI device) that would let you know when your visual system is *processing* infrared light (even though you would have no conscious sensation, no quale, of it). By hypothesis, you would have no quale of this, but perhaps it would nevertheless feel like something, perhaps like an intuition ("I don't sense any color or light, but somehow I just feel that something's different when I'm exposed to infrared light"). This would be a quale, I suppose, but not necessarily a quale *of* the infrared light. (Maybe it would be a second-order quale?) But it would probably not be like a visual experience of, say, red light; it wouldn't be, say, a "deeper" red.

So, what would be the difference in my behavior between my reaction to a quale of red light and my reaction to quale-less infrared light? In the situation where I am aware of a red quale, I can *voluntarily* react (or not react) to the quale of red light: I can decide to stop my car, or I can decide to try to run the light if there's no traffic or pedestrians. This freedom of choice does not run afoul of the problem of free will, because whatever free will turns out to be will work here: Either there really is such a thing as free will, in which case I have the free will to react or not to react to the red quale, or there really is no such thing as free will, in which case whatever explanation one has of the *appearance* of free will will be an explanation of my decision.

However, in the situation where I am *not* aware of an infrared quale (and even more so if there is not even an "intuition" of a sensation), I would have no choice in the matter: If the infrared light causes me to do something or to react in some way, I would do it, or react, without having the opportunity to choose not to. If a certain unsmelled pheromone causes me to be attracted to someone, I will be attracted; I will have no choice (or no "choice", in case there is no free will) in the matter.

Thus, we *can* be *partial* zombies but not *complete* zombies, because a complete zombie would have no free will, whereas we do. If a complete zombie had free will, it would then also have to have *some* higher-order thought, or access, or awareness of the impingement of the external object (chemicals for odor, photons for vision, etc.) on its sensory apparatus. But then, perhaps, it wouldn't be a zombie.

Knowing that the external object has thus impinged—being *aware* of its impingement—*from the first-person point of view* (i.e., subjective awareness, not objective awareness) *is* the experience of a quale. But there is still the possibility that your quale might be "inverted" with respect to mine. I once suggested that this might be due to (physical) implementation differences between us (Rapaport, 2005b). We will explore this further in what follows.

Why are some stimuli experienced as colors and others as sounds? Could this be because of their different sources? After all, photons produce color experiences, sound waves produce sounds, airborne chemicals produce odors. Or could it be because of the different sense organs that receive the photons, sound waves, or chemicals? But they all transduce these varied stimuli into a *single* neural "language of thought" (Piccinini, 2020, pp. 213–214). So these *could* all be experienced on a common spectrum rather than by orthogonal experiences. (Perhaps they are, for those of us who experience synaesthesia or "cross-modal reorganization" (Allman et al., 2009).)

For that matter, why are some things experienced at all (and some of those experienced differently), but others are not? Perhaps this is because of evolutionary usefulness: our voluntary ability to perform the so-called four Fs (feeding, fleeing, fighting, and reproduction); after all, odors are not as useful to us as they are to dogs, so it makes sense that we might not experience as many of them.

Chalmers (1996) observes that a mental state or process can be characterized in two different ways: functionally and "intrinsically": A functional characterization would be "external" to the mental state or process, characterizing it in terms of its inputs, its relations to other mental states and processes, and its outputs. In short, it would be a characterization in terms of its "role in . . . [the] cognitive economy" of the cognitive agent (Chalmers, 1996, p. 16). Almost by definition, it would be a characterization independent of its implementation. It would, perhaps, be Wittgenstein's beetle-box (with or without a beetle).

An intrinsic characterization would be "internal" or "phenomenal", a characterization in terms of "what it's like"—its quale. Almost by definition, it would be a characterization independent of its functional role. It would be a beetle in the box.

This notion of a cognitive economy is worth pursuing. Compare the world's *monetary* economy: A dollar has the value it does because of the *role* that it plays in the world monetary economy.[6] Does it have an *intrinsic* value? (Does it have a "quale"?) Dennett (2001) asks whether a dollar has "something logically independent of its functionalistic exchange powers" and answers that it does not, that

---

[6]Although a currency might play several roles at once; cf. Hofstadter 1997, Ch. 15, on the incompatible roles of the Polish złoty.

there is no economic theory of such an intrinsic value. But one possibility for its intrinsic value is the value of the paper it's printed on. (Of course, in the case of a dollar, that value is itself (recursively) a function of the world monetary economy!)

Is there an analogous way to pin down an intrinsic characterization of a quale?

6 Numbers

Let's turn to a seemingly unrelated question. Is there any way to pin down the intended interpretation of an axiom system, say, Peano's axioms for the natural numbers?

We can characterize the natural numbers only as: any sequence that satisfies Peano's axioms (an "ω-sequence"). But an infinite variety of sequences do that. There is no way to pick out "the" natural numbers. Trying to do that would be like trying to characterize qualia! In Quine's words (1969: 45), "Arithmetic is ... all there is to number: there is no saying absolutely what the numbers are; there is only arithmetic". Arithmetic, like the cognitive economy, can only be characterized functionally, or—to use the mathematical term—*structurally*. Numbers, like qualia, are ineffable. But whereof we cannot speak, *must* we thereof be silent?

Paul Benacerraf (1965) argued that numbers could not be sets. The number 3 is neither "Ernie" 's $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ nor "Johnny" 's $\{\{\{\emptyset\}\}\}$ (Benacerraf, 1965, p. 55)—and it certainly cannot be both: Each of those sets has properties and relations that the other lacks *and* that are irrelevant to numbers. For example, on both views, not only do numbers wind up having set-theoretical cardinality (which is irrelevant),[7] but the same number would have different cardinalities: On the first view, the cardinality of 3 is 3; on the other, it is 1. And on the first view, $1 = \{\emptyset\} \in 3$, whereas, on the other, $1 = \{\emptyset\} \notin 3$; both cannot be true. And *none* of these are true of the natural number 3 solely as characterized by Peano's axioms. As Benacerraf (1965, p. 70) says, "*Any* object can *play the role of* 3".

Although either one (but not both together) of these views of the number 3 gives us something that the number 3 *could be*, they do not tell us what it *is*. Compare the similar situation of ordered pairs:

> The ordered pair (a, b) is defined to be $\{\{a\}, \{a, b\}\}$, and from this definition the basic law of ordered pairs, that $(a, b) = (c, d)$ iff $a = c$ and $b = d$, is deduced. *It is not pretended that this definition reveals what ordered pairs 'really were all along.'* What the definition and derivation of the basic law do show is that the positing of ordered

---

[7] Well, almost: The cardinality of $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ is designed to match the fact that '3' represents the cardinal number 3.

pairs subject to this basic law as entities over and above sets is, in a sense, superfluous. (Burgess, 2001, p. 62, my emphasis)

Are qualia similarly superfluous?[8]

Numbers are in a similar situation:

Arithmetic is ... the science that elaborates the abstract sructure that all progressions have in common. ... It is not ... concerned with particular objects—the numbers. The search for which independently identifiable particular objects the numbers really are ... is ... misguided ... . (Benacerraf, 1965, p. 70)

Is the search for an intrinsic characterization of qualia, ineffable though they may be, similarly misguided?

Let us examine these analogies more closely. What is a graph? It can be defined as (a structure) consisting of a set V of vertices and a set E of edges, with certain relationships among the members of V and E. But what is a "vertex"? What is an "edge"? Answer: anything that satisfies those relationships. So, a telephone network is not merely a physical system that can be *modeled as* a graph; it really *is* a graph, because we can take phones to be vertices and phone connections to be edges. Logically speaking, vertices and edges are (types of) variables that can take as values certain phones and phone connections. Such talk of variables is just talk of roles that can be played by certain (usually physical) objects. The objects *implement* those roles. The roles are *abstractions* (Rapaport, 1999, 2005b).

Oswald Veblen (1904, p. 344) made this point in his axiomatization of geometry: "The terms 'point' and 'order' ... differ from the other terms of geometry in that they are *undefined*." Because they are undefined, we are not told *what they are*; we are only told how they relate to other geometrical objects—i.e., we are only told what their roles are in the functional "economy" of geometry. Therefore, they can "*be*" (better: be implemented by) anything that can play those roles. Hilbert expressed this same idea more famously: "One must be able to say at all times—instead of points, lines, and planes—tables, chairs, and beer mugs".[9]

---

[8]However, Burgess can be read as suggesting that it is the beetle-*box* that is superfluous. If you have a beetle, perhaps you don't need to have a box to keep it in. And if you lack a beetle, you certainly don't need a box.

[9]Hilbert's observation appears in his *Gesammelte Abhandlungen* ("Complete Works"), vol. 3, p. 403, as cited in Coffa 1991, p. 135; see also Stewart Shapiro 2009, p. 176. Elsewhere, Hilbert used a different example:

... it is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of in any way one likes. If in speaking of my points, I think of some system of things, e.g., the system: love, law, chimney-sweep ... and then assume all

But Veblen (1904, p. 346) also said that "there is *essentially only one* class of which the … axioms are valid", i.e., one class "up to isomorphism". Veblen continues: "In more exact language, any *two* classes K and K' of objects that satisfy the … axioms are capable of a one-to-one correspondence" between them; i.e., they are isomorphic. Hilbert would agree, but a set of points and lines (as Euclid thought of them) that satisfy the axioms, and a set of tables and chairs that satisfy the axioms, are like qualia and inverted qualia.

Nicholas P. White (1974) tried to rebut Benacerraf, but their theories are really consistent: White focused on the role-*filler* rather than the *role*. He suggested that a certain set (e.g., $\{\{\{\emptyset\}\}\}$) is a 3 in a certain series (p. 113); i.e., it plays the role of a 3, just as Richard Burton played the role of Hamlet in a certain 1964 production of *Hamlet*. "There are indeed numbers, and there are plenty of them" (p. 118); yes, and there are plenty of different qualia, too. White's way of looking at things focuses on the implementation; Benaceraff's focuses on the functional economy. Their views are complementary, not inconsistent.

Elsewhere, I have urged that the number 3 is anything that *implements* the third item in a sequence that satisfies the *abstraction* described by Peano's axioms (Rapaport, 1999). Now I urge that a quale of a mental state or process is anything that implements that mental state or process, where the mental state or process is characterized functionally.

There are two ways to interpret answers to a question of the form "X is Y" (cf. Gert (2008)): On one interpretation, to say that X is Y is to explicate X in terms of a functional theory describing what properties any candidate for X must satisfy, by giving the "script" for any actor playing the role of X. Thus, "3 is the third item in a Peano-satisfying sequence", or "water is the clear, odorless, tasteless liquid found in rivers and oceans that we use for drinking and washing", or "Hamlet is a prince of Denmark who can't decide whether to avenge his father's murder" would be such statements. On another interpretation, to say that X is Y is to give a particular "actor" that plays that role. Thus, "3 is $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$", or "water is H_2O", or "Hamlet is Burton" would be statements of this kind. It might be better always to make the context explicit: *In Ernie's theory,* 3 is $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$. *On Earth,* water is H_2O. *In the 1964 Broadway production,* Hamlet is Burton.[10]

Consider a functional characterization of water, e.g., my characterization of the "narrow" meaning of 'water' as given in the previous paragraph. Then H_2O is just

my axioms as relations between these things, then my propositions, e.g., Pythagoras' theorem, are also valid for these things … [A]ny theory can always be applied to infinitely many systems of basic elements. (Quoted in Stewart Shapiro 2013, p. 168.)

---

[10]Compare the similar solution to problems of fictional characters; see, e.g., Lewis 1978; Rapaport 1991; Rapaport and Shapiro 1995.

one "actor" that can play that role, and Twin Earth's XYZ is another. Or, to switch examples, it's not that Hamlet *is* Richard Burton (or vice versa); rather, Burton is one among many who have played the role. It is only in a particular production that we can say that Hamlet is Burton. More precisely, the *role* of Hamlet was *played* by Burton. Hamlet and Burton are not identical, after all. Just as $\{\{\{\emptyset\}\}\}$ is 3, in the sense of "is a 3" or "plays the role of 3", so we can say that Burton is Hamlet, in the sense that Burton plays the role of Hamlet. Compare these to "3 = the successor of 2", which is a functional or structural definition.

To sum up: The structural view of mathematics neither requires nor allows us to specify *what* 3 *is*. It only defines 3 in terms of its role in arithmetic. Still, when I do arithmetic, I implement 3 somehow, perhaps as $\{\{\{\emptyset\}\}\}$, or perhaps as the numeral '3' (most likely, especially when writing), or perhaps as an internal, mental "numeron".[11]

7 Qualia

7.1 Neural and Computational Implementations

Similarly, the structural or functional view of qualia neither requires nor allows us to specify *what* qualia *are*. It only defines them in terms of their role in the cognitive economy. Still, they have to be implemented (if they are not absent), and thus I *do* experience red in a certain way.

As noted earlier (§2), Piccinini (2020, p. 182) argues against a traditional, two-level view of the mental-neural relationship. On this view, an abstract, functional analysis of mental phenomena is independent of a neural, implementation-level analysis, primarily because the former can be multiply realized and are physical-medium independent. Piccinini (pp. 166, 192) promotes an "integrated" view that requires both levels, each of which constrains the other (p. 194), and neither of which is independent of the other. If the role that qualia play in the cognitive economy is to provide qualitative ("feel") properties that accompany (i.e., that play the role of) functional properties (p. 246), then they are part of the implementation level of an integrated, "mechanistic" account of (human) cognition.

Certainly, both levels may be needed to explain *human* cognition (a cognitive-science task). But what about AI, viewed as a computational theory of cognition that might be applicable to non-human agents? Can a cognitive computer or robot have qualia? I see no reason why not, though theirs might differ from ours (as in the inverted spectrum). On this view, although qualia might not be purely mental

---

[11]"Numerons are any distinct and arbitrary tags that a mind (human or nonhuman) uses in enumerating a set of objects" (Gelman and Gallistel, 1986, p. 77).

(despite such having such mental properties as privacy), they are, or can be, an essential part of a cognitive neuroscientific theory of mental phenomena just as well as a computational one.

Others agree:

Stanislaus Dehaene and Lionel Naccache (2001, p. 30) say that qualia might be "biological properties" of consciousness's "workspace":

> each workspace state is "highly differentiated" and of "high complexity". ... Thus the flux of neuronal workspace states associated with a perceptual experience is vastly beyond accurate verbal description or long-term memory storage .... [A]lthough the major organization of this repertoire is shared by all members of the species, its details result from a developmental process of epigenesis and are therefore specific to each individual. Thus, the contents of perceptual awareness are complex, dynamic, multi-faceted neural states that cannot be memorized or transmitted to others in their entirety.

Similarly, qualia are ineffably "beyond accurate verbal description", and it is quite possible that yours are quite different from mine; this accounts for inverted (if not absent) qualia.

Block (2001, pp. 203) says that functionalism "identifies consciousness with a role"—just like 3 = the successor of 2 or water = clear liquid, etc. But "physicalism identifies consciousness with a physical or biological property that *fills or implements or realizes* that role in humans"—just like 3 can be taken as $\{\{\{\emptyset\}\}\}$ or water as H_2O. Commenting on views like that of Dehaene and Naccache, Block (2001, p. 204) asks: "How do you know that it is **broadcasting in the global workspace** that makes a representation conscious as opposed to something about the *human biological realization* of that broadcasting that makes it conscious?" (my boldface, Block's italics). My answer: You don't! Hence the possibility of absent qualia (and zombies).

Stuart C. Shapiro and Jonathan P. Bona (2010) distinguish among a "knowledge layer" (KL), a "perceptuo-motor layer" (PML), and a "sensori-actuator layer" (SAL) in their architecture for an implemented computational cognitive agent (named 'Cassie'). The KL is software that corresponds to the "mind" of the agent, the SAL is entirely hardware that corresponds to its "body",[12] and "The PML comprises three increasingly-abstract sub-layers, which connect to the SAL at one end and the KL at the other. ... the PML layers [can be] implemented in software" (Bona, personal communication, 4 December 2021).

---

[12] According to Bona (personal communication, 4 December 2021).

There are KL terms for every mental entity Cassie has conceived of, including ... colors .... There are PML structures ... for features of the perceivable world that Cassie's perceptual apparatus can detect and distinguish. (Shapiro and Bona, 2010, §6.1)

They identify qualia with "PML-descriptions" (Shapiro and Bona, 2010, §6.7), which are *n*-tuples of "feature values" some of which can be null (i.e., "absent").[13] PML-descriptions are much like role-fillers or beetles in boxes:

Each KL term for a perceivable entity, category, or property is grounded by being aligned with a PML-description, possibly with unfilled (null) components. For example, [one particular version of] Cassie used two-component PML-descriptions in which the domains were color and shape. The KL term denoting Cassie's idea of blue was aligned with a PML-description whose color component was the PML structure the vision system used when it detected blue in the visual field .... (Shapiro and Bona, 2010, §6.1)

7.2 Ineffability

As Galen Strawson (2021, p. 239) points out, Bertrand Russell held that physics only gives us *structural* knowledge of the world, but "We know nothing of the intrinsic quality of physical phenomena" (Russell, 1927, p. 154). We have knowledge by *description* of the *structure* "of physical phenomena", but our experience of qualia is knowledge by *acquaintance* of its "intrinsic quality".

Qualia are (physically) ineffable; we cannot describe them. Consider a famous person whom I have not met, say, Paul McCartney. I can know *who* he is by description, but that only gives me abstract, structural information about him. Now consider a famous person whom I *have* met, say, the singer and actress Lucie Arnaz. Before I met her, I only had abstract, structural, descriptive knowledge of her. Whenever I have met her in person, I had concrete knowledge by acquaintance of her that I cannot convey in language, for any attempt to do that would just be more (abstract, structural) description. The "content" of my knowledge by acquaintance of her is "incommunicable"; it is only experienceable, by me.

Why is a quale not describable? Prima facie, it is: Suppose that we are on opposite sides of a fire engine, so that neither of us can see each other. Can't I describe the red color of my shirt by telling you that it is just like the red color of the fire engine that we are both looking at? Yes, but that only gives you a *relational*

---

[13] However, Bona (2013) says "One might claim that the agent has a private qualitative experience of the motion of its wheels that might very well constitute something like qualia—but I make no such claim here."

idea of the color that I am experiencing when I look at my shirt. And if your quale of the fire engine's red differs from mine (which it can; Block 1999), then all you can know about *my* quale (*my* experience) of my shirt is that I say that it is like yours of the fire engine. The same is true of oenophiles' descriptions of the flavor of a wine. They can't *tell* you what the flavor *is*; they can only tell you what it is similar to ("what it is like", to use an overused expression), as in the following example:

> Aromas of peach and nectarine and juicy melon mingle with hints of vanilla, toasted nuts, and light caramel. The creamy, rich palate bursts with ripe pear and peach leading to a long, lush finish. (https://www. sonomacutrer.com/wine/simply-cutrer/)

If you don't know the flavor of peaches, etc., then you won't know the flavor of this wine from this description of it. If you don't know the value of either term in an equivalence or an analogy, you can't fully understand it: Typically, one understands one of them in terms of the other, which must be antecedently understood (Rapaport, 1995). Perhaps the same holds for the relation between a mental state or process and its quale: By experiencing (hence understanding) the quale, we thereby come to understand (and undergo) the mental state or process. The only way for you to know what the flavor *is* is to taste it (but then, of course, your taste quale might not be the same as mine, even if both are like peach, nectarine, melon, etc.). Descriptions are just structural information. The best that I can do to describe my quale is to talk "around" it, to talk "about" it; I cannot get you inside my head to experience it. Whereof we cannot speak, thereof we can only express analogies.

7.3 Problems

Although I am perfectly content with the inability to uniquely characterize the natural numbers, I am discontent with the inability to uniquely characterize qualia. Are natural numbers and qualia not analogous?

Possibly, the quale of a mental state or process is not characterizable in terms of its functional relationships but only in terms of how the experiencer relates to it. But that runs into the value-of-a-dollar problem: It's ultimately defined in terms of other mental states and processes, which leads us back to Dennett.

Perhaps the nature of a quale must lead *out* of the network of mental states and processes into the physical interpretation. Then we would have to allow for inverted and absent qualia, depending on the implementation. This is a bullet we might have to bite! It gets us out of the mental-functional circle, and it remains physical. (See Rapaport 2005.)

Consider a recursively defined function, with a base case $f(0)$ and a recursive case $f(n+1)$:

$$f(0) = q;$$
$$f(n+1) = g(f(n))$$

The recursive case is like a functional theory of access consciousness, and the base case is like the quale of phenomenal consciousness. The base case is a particular implementation of the recursion. If we have another recursive function with a different base case but the same recursive case:

$$h(0) = r, \text{ where } r \neq q;$$
$$h(n+1) = g(h(n))$$

then we have two distinct functions with the same "functional" theory (the recursive case) but different "qualia" (different base cases). And recursive cases without base cases are like absent qualia.[14]

If mental states and processes are physically implemented, then qualia *are* part of the big picture, after all. They are values of variables. Those *variables* are part of the mental (functional, computational) theory; their *values* are part of the physical implementation of that theory: They are a side effect, an implementation-level detail. Qualia-variables without values are "absent qualia" (such mental states or processes would be unconscious); qualia-variables with different values are "inverted qualia".

8   Summary and Conclusion

Let us return to our triad. Rather than talking about qualia *simpliciter*, I suggest that we distinguish between qualia-*roles* and qualia-*fillers*: There are roles that qualia play, and there are fillers of those roles. Qualia-roles are mental and functional; qualia-fillers are not.

Where we draw the dividing line between the mental and the physical is arbitrary: If we want to maintain that qualia-fillers *are* mental, then we must accept that proposition 1 is false: The mental is not (solely) functional. But that's OK; it's consistent with Piccinini's single-level, functional+neural analysis. On the other hand, if we want to maintain that qualia-fillers are *not* mental (because not functional; i.e., that proposition 3 is true), then that's OK, too, and equally consistent with Piccinini's analysis. If qualia-fillers are not functional, then they don't *have* to be computational; but they can be.

---

[14]Perhaps Morbini and Schubert 2005 can be read in this way.

18

Qualia-roles, being functional, can be implemented as variables in a computational theory; they are beetle-boxes. Qualia-fillers would be the values of those variables; they are beetles. And just as a computer might do arithmetic as Benacerraf's Ernie does, rather than as his Johnny does, so a (computational) cognitive agent might experience a red quale-role via a green quale-filler rather than via a red one.

A final word: Although I think that distinguishing between mental-functional qualia-roles and physical qualia-fillers is a good way to think about things, it does not *solve* the hard problems of consciousness. For even if I experience red because my qualia-filler (my neuron firings) is of the red variety rather than the green variety, I still don't know why it is that I experience it *as* red, or why I don't experience it as the *sound* of a bell, or—for that matter—why (or how) I *experience* it at all. There's still a lot of work to be done.

9   Acknowledgment

References

Allman, B. L., L. P. Keniston, and M. A. Meredith (2009, 7 April). Adult deafness induces somatosensory conversion of ferret auditory cortex. *PNAS 106*(14), 5925–5930. https://www.pnas.org/content/106/14/5925.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Angier, N. (1992, 30 January). Odor receptors discovered in sperm cells. *New York Times*, A19. https://www.nytimes.com/1992/01/30/us/odor-receptors-discovered-in-sperm-cells.html.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge, UK: Cambridge University Press.

Baars, B. J. and S. Franklin (2009, June). Consciousness is computational: The LIDA model of global workspace theory. *International Journal of Machine Consciousness 1*(1), 23–32. http://ccrg.cs.memphis.edu/assets/papers/2009/GWT-IJMC-2009.pdf.

Benacerraf, P. (1965, January). What numbers could not be. *Philosophical Review 74*(1), 47–73.

Block, N. (1978). Troubles with functionalism. In C. Savage (Ed.), *Minnesota Studies in the Philosophy of Science, Volume 9*, pp. 261–325. Minneapolis: University of Minnesota Press. http://mcps.umn.edu/philosophy/9_12Block.pdf.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences 18*, 227–287. http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/1995_Function.pdf.

Block, N. (1999, Spring/Fall). Sexism, racism, ageism, and the nature of consciousness. *Philosophical Topics 26*(1/2), 39–70. http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/1999_Sexism.pdf.

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition 79*, 197–219. https://www.nyu.edu/gsas/dept/philo/faculty/block/papers/Elsevier.html.

Bona, J. P. (2013, 14 August). MGLAIR: A multimodal cognitive agent architecture. https://www.cse.buffalo.edu/sneps/Bibliography/bona-dissertation.pdf. PhD dissertation, Department of Computer Science & Engineering, State University of New York at Buffalo.

Burgess, J. P. (2001). Set theory. In L. Goble (Ed.), *The Blackwell Guide to Philosophical Logic*, pp. 55–71. Malden, MA: Blackwell.

Calef, S. (nd). Dualism and mind. In J. Fieser and B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/dualism/.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies 2*(3), 200–219. http://consc.net/papers/facing.html.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press. http://tinyurl.com/plv877.

Chen, D. and J. Haviland-Jones (2000). Human olfactory communication of emotion. *Perceptual and Motor Skills 91*, 771–781.

Chirimuuta, M. (2021). Your brain is like a computer: Function, analogy, simplification. In F. Calzavarini and M. Viola (Eds.), *Neural Mechanisms: New Challenges in the Philosophy of Neuroscience*, pp. 235–261. Cham, Switzerland: Springer. https://ruccs.rutgers.edu/images/talks-materials/Chirimuuta_Brain-Computer-Analogy_Rutgers.pdf.

Coffa, J. (1991). *The Semantic Tradition from Kant to Carnap: To the Vienna Station*. Cambridge, UK: Cambridge University Press.

Crick, F. (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Simon & Schuster.

Dehaene, S. and L. Naccache (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition 79*, 1–37. https://www.nyu.edu/gsas/dept/philo/courses/consciousness05/Dehaene1.pdf.

Dennett, D. C. (1975). Why the law of effect will not go away. *Journal for the Theory of Social Behaviour 5*(2), 169–188. Reprinted in Daniel C. Dennett, *Brainstorms* (Montgomery, VT: Bradford Books): 71-89, https://dl.tufts.edu/downloads/j9602b715?filename=2r36v862p.pdf.

Dennett, D. C. (1988). Quining qualia. In A. Marcel and E. Bisiach (Eds.), *Consciousness in Modern Science*. Oxford: Oxford University Press. https://ase.tufts.edu/cogstud/dennett/papers/quinqual.htm.

Dennett, D. C. (1992). *Consciousness Explained*. Boston: Little, Brown and Co.

Dennett, D. C. (2001, 12 December). Consciousness: How much is that in *real* money? https://ase.tufts.edu/cogstud/dennett/papers/consciousnessmoney.htm.

Dennett, D. C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Fodor, J. A. (1968, October). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy 65*(20), 627–640. https://havingthought2012.files.wordpress.com/2012/01/fodor-the-appeal-to-tacit-knowledge-in-psychological-explanation.pdf.

Fountain, H. (2006, 3 October). This plant has the sense of smell (loves tomatoes, hates wheat). *New York Times*, F1. https://www.nytimes.com/2006/10/03/science/03find.html.

Gelman, R. and C. Gallistel (1986). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

Gert, J. (2008, March). What colors could not be: An argument for color primitivism. *Journal of Philosophy 105*(3), 128–155.

Greenwood, V. (2021, 4 March). How do blind worms see the color blue? *New York Times*. https://www.nytimes.com/2021/03/04/science/worms-colors-blind.html.

Hofstadter, D. R. (1997). *Le Ton beau de Marot: In Praise of the Music of Language*. New York: BasicBooks.

Jackson, F. (1982, April). Epiphenomenal qualia. *Philosophical Quarterly 32*(127), 127–136.

Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy 83*, 291–295.

Jackson, F. (1995). Postscript on 'What Mary didn't know'. In P. Moser and J. Trout (Eds.), *Contemporary Materialism*, pp. 184–189. London: Routledge.

Jacquette, D. (1996, May). What would a cerebroscope do? *Journal of the British Society for Phenomenology 27*(2), 188–199. https://ur.booksc.eu/dl/62843039/cb4b72.

Kirk, R. (1974, January). Sentience and behaviour. *Mind 83*(329), 43–60.

Kriegel, U. (2009). *Subjective Consciousness: A Self-Representational Theory*. New York: Oxford University Press.

Levin, J. (2021). Functionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/functionalism/.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly 64*, 354–361. https://www.newdualism.org/papers/J.Levine/Levine-PPQ1983.pdf.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly 15*, 37–46.

Lycan, W. G. (1981, January). Form, function, and feel. *Journal of Philosophy 78*(1), 24–50.

Maley, C. and G. Piccinini (2013). Get the latest upgrade: Functionalism 6.3.1. *Philosophia Scientiae 17*(2), 135–149. https://journals.openedition.org/philosophiascientiae/pdf/861.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W.H. Freeman.

McGinn, C. (2021, 22 March). Qualia. https://www.colinmcginn.net/qualia/.

Mehta, N. (2022, January). The fragmentation of phenomenal character. *Philosophy and Phenomenological Research 104*(1), 209–231.

Mills, H. D. (1971). Top-down programming in large systems. In R. Rustin (Ed.), *Debugging Techniques in Large Systems*, pp. 41–55. Englewood Cliffs, NJ: Prentice-Hall.

Morbini, F. and L. K. Schubert (2005, September). Conscious agents. Technical Report TR879, University of Rochester Computer Science Department. https://urresearch.rochester.edu/fileDownloadForInstitutionalItem.action?itemId=1893&itemFileId=2452.

Nagel, T. (1974, October). What is it like to be a bat? *Philosophical Review 83*(4), 435–450. https://www.sas.upenn.edu/~cavitch/pdf-library/Nagel_Bat.pdf.

Pattis, R. E., J. Roberts, and M. Stehlik (1995). *Karel the Robot: A Gentle Introduction to the Art of Programming, Second Edition*. New York: John Wiley & Sons.

Piccinini, G. (2020). *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford: Oxford University Press.

Pylyshyn, Z. (2003, March). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences 7*(3), 113–118. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.3609&rep=rep1&type=pdf.

Rapaport, W. J. (1985). Machine understanding and data abstraction in Searle's Chinese room. In *Proceedings of the 7th Annual Conference of the Cognitive Science Society*, pp. 341–345. Hillsdale, NJ: Erlbaum. http://www.cse.buffalo.edu/~rapaport/Papers/rapaport85-cogscisoc.pdf.

Rapaport, W. J. (1986, Summer). Philosophy, Artificial Intelligence, and the Chinese-room argument. *Abacus: The Magazine for the Computer Professional 3*, 6–17. Correspondence, *Abacus* 4 (Winter 1987): 6–7; 4 (Spring): 5–7; http://www.cse.buffalo.edu/~rapaport/Papers/abacus.pdf.

Rapaport, W. J. (1988, December). To think or not to think: Review of Searle's *Minds, Brains and Science*. *Noûs 22*(4), 585–609. http://www.cse.buffalo.edu/~rapaport/Papers/2Tor-2T.pdf.

Rapaport, W. J. (1991). Predication, fiction, and artificial intelligence. *Topoi 10*, 79–111.

Rapaport, W. J. (1995). Understanding understanding: Syntactic semantics and computational cognition. In J. E. Tomberlin (Ed.), *AI, Connectionism, and Philosophical Psychology (Philosophical Perspectives, Vol. 9)*, pp. 49–88. Atascadero, CA: Ridgeview. http://www.cse.buffalo.edu/~rapaport/Papers/

rapaport95-uu.pdf. Reprinted in Toribio, Josefa, & Clark, Andy (eds.) (1998), *Language and Meaning in Cognitive Science: Cognitive Issues and Semantic Theory (Artificial Intelligence and Cognitive Science: Conceptual Issues*, Vol. 4) (New York: Garland).

Rapaport, W. J. (1998). How minds can be computational systems. *Journal of Experimental and Theoretical Artificial Intelligence 10*, 403–419. http://www.cse.buffalo.edu/~rapaport/Papers/jetai-sspp98.pdf.

Rapaport, W. J. (1999). Implementation is semantic interpretation. *The Monist 82*, 109–130. http://www.cse.buffalo.edu/~rapaport/Papers/monist.pdf.

Rapaport, W. J. (2000, October). How to pass a Turing test: Syntactic semantics, natural-language understanding, and first-person cognition. *Journal of Logic, Language, and Information 9*(4), 467–490. http://www.cse.buffalo.edu/~rapaport/Papers/TURING.pdf. Reprinted in James H. Moor, *The Turing Test: The Elusive Standard of Artificial Intelligence* (Dordrecht, The Netherlands: Kluwer Academic, 2003): 161–184.

Rapaport, W. J. (2005a). Castañeda, Hector-Neri. In J. R. Shook (Ed.), *The Dictionary of Modern American Philosophers, 1860–1960*, pp. 452–412. Bristol, UK: Thoemmes Press. http://www.cse.buffalo.edu/~rapaport/Papers/hncdict.tr.pdf.

Rapaport, W. J. (2005b, December). Implementation is semantic interpretation: Further thoughts. *Journal of Experimental and Theoretical Artificial Intelligence 17*(4), 385–417. https://www.cse.buffalo.edu//~rapaport/Papers/jetai05.pdf.

Rapaport, W. J. (2012, January-June). Semiotic systems, computers, and the mind: How cognition could be computing. *International Journal of Signs and Semiotic Systems 2*(1), 32–71. http://www.cse.buffalo.edu/~rapaport/Papers/Semiotic_Systems,_Computers,_and_the_Mind.pdf. Revised version published as Rapaport 2018.

Rapaport, W. J. (2018). Syntactic semantics and the proper treatment of computationalism. In M. Danesi (Ed.), *Empirical Research on Semiotics and Visual Rhetoric*, pp. 128–176. Hershey, PA: IGI Global. References on pp. 273–307; http://www.cse.buffalo.edu/~rapaport/Papers/SynSemProperTrtmtCompnlism.pdf. Revised version of Rapaport 2012.

Rapaport, W. J. (2023). *Philosophy of Computer Science: An Introduction to the Issues and the Literature*. Wiley.

Rapaport, W. J. and S. C. Shapiro (1995). Cognition and fiction. In J. F. Duchan, G. A. Bruder, and L. E. Hewitt (Eds.), *Deixis in Narrative: A Cognitive Science Perspective*, pp. 107–128. Hillsdale, NJ: Lawrence Erlbaum Associates. http://www.cse.buffalo.edu/~rapaport/Papers/rapaport.shapiro.95.cogandfict.pdf.

Rosenthal, D. M. (1986, May). Two concepts of consciousness. *Philosophical Studies 49*(3), 329–359. https://www.davidrosenthal.org/DR-TCC.pdf.

Rosenthal, D. M. (2009). Higher-order theories of consciousness. In B. McLaughlin and A. Beckermann (Eds.), *Oxford Handbook on the Philosophy of Mind*, pp. 239–252. Oxford: Clarendon Press. https://www.davidrosenthal.org/DR-HO-Theories-Handbook.pdf.

Russell, B. (1927). *An Outline of Philosophy*. London: George Allen and Unwin.

Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Shagrir, O. (1997). Two dogmas of computationalism. *Minds and Machines 7*, 321–344. https://link.springer.com/content/pdf/10.1023/A:1008236522699.pdf.

Shapiro, S. (2009, March). We hold these truths to be self-evident: But what do we mean by that? *Review of Symbolic Logic 2*(1), 175–207. http://www.pgrim.org/philosophersannual/29articles/shapirowehold.pdf.

Shapiro, S. (2013). The open texture of computability. In B. J. Copeland, C. J. Posy, and O. Shagrir (Eds.), *Computability: Turing, Gödel, Church, and Beyond*, pp. 153–181. Cambridge, MA: MIT Press.

Shapiro, S. C. and J. P. Bona (2010). The GLAIR cognitive architecture. *International Journal of Machine Consciousness 2*(2), 307–332. https://www.cse.buffalo.edu/~shapiro/Papers/glairPublished.pdf.

Simon, H. A. (1996). *The Sciences of the Artificial, Third Edition*. Cambridge, MA: MIT Press.

Sobel, N., V. Prabhakaran, C. A. Hartley, J. E. Desmond, G. H. Glover, E. V. Sullivan, and J. D. Gabrieli (1999). Blind smell: Brain activation induced by an undetected air-borne chemical. *Brain 122*, 209–217.

Strawson, G. (2012). Real naturalism. *Proceedings and Addresses of the American Philosophical Association 86*(2), 125–154.

Strawson, G. (2021). 'Oh you materialist!'. *Journal of Consciousness Studies 28*(9–10), 229–249. https://philpapers.org/archive/STROYM.pdf.

Tye, M. (2021). Qualia. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/qualia/.

Veblen, O. (1904). A system of axioms for geometry. *Transactions of the American Mathematical Society 5*(3), 343–384. https://www.ams.org/journals/tran/1904-005-03/S0002-9947-1904-1500678-X/S0002-9947-1904-1500678-X.pdf.

White, N. P. (1974). What numbers are. *Synthese 27*, 111–124.

Wirth, N. (1971, April). Program development by stepwise refinement. *Communications of the ACM 14*(4), 221–227. http://oberoncore.ru/_media/library/wirth_program_development_by_stepwise_refinement2.pdf, http://sunnyday.mit.edu/16.355/wirth-refinement.html, and http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.9376&rep=rep1&type=pdf.

Wittgenstein, L. (1958). *Philosophical Investigations, Third Edition*. New York: Macmillan.