

Doublemindedness: A model for a dual content cognitive architecture

Jenann Ismael
Centre for Time, University of Sydney
University of Arizona
jtismael@u.arizona.edu

Introduction

There has been a confluence of interest in recent years in self-representational accounts of phenomenal consciousness. These accounts developed out of the higher order thought (HOT) theory of David Rosenthal and have been held to exhibit virtues of that account without pitfalls. There is historical precedent for thinking that self-representation is somehow central to consciousness in the views of Aristotle, Kant, Locke, and the phenomenological tradition, and some quite interesting work has been done to answer charges that have been lodged against the self-representational view. But the outstanding stumbling blocks to any reductive account of phenomenal consciousness remain the subjectivity of phenomenal properties and cognitive and epistemic gaps that plague the relationship between physical and phenomenal properties.¹ I'll show how a self-representational account elucidates subjectivity and explains the source of those gaps, and I'll close with a few remarks about self-representational accounts that aim to provide analyses of what it is to be conscious.

The dual content structure

Imagine a small universe – say, a room with a fixed number of inhabitants – and suppose you wanted to design a filing cabinet to store information about the room. Suppose that the cabinet – call it F – keeps files on everything, cataloguing the properties of the objects to which they pertain. Perhaps the room contains other cabinets, each with its own set of files, all faithfully catalogued in F. Since F is in the room, it contains a file on itself, but when it comes to cataloguing its own contents, instead of attempting to reproduce the entirety of its structure in a single file, it employs a simple expedient. It leaves a little note in the file labeled 'F' that says 'see cabinet'. Under this arrangement, every file *in* F does double representational duty; it has the significance it always did of representing some object in the cabinet's external environment, but it also now represents something internal to the cabinet: *viz.*, itself. The note effectively turns the cabinet into a model of itself, imposing a new, secondary interpretation on contents that already had a representational significance. Under the ordinary, intentional interpretation, files in F (which I'll treat, for convenience, as the smallest semantically significant components of F) represent elements in the room. Under the new

¹ What is meant by 'subjectivity' is not often clearly defined. It's one of the virtues of the dual content model that the notion can be given a definite sense in that context.

interpretation, files in F represent themselves.² The arrangement circumvents the impossibility of containing an isomorphic copy of oneself as a proper part, avoids unnecessary reification of structure by making use of structure that is already present in the cabinet, and has the virtue of unassailable accuracy.

How do we transpose this picture to ourselves? We need only some very general assumptions about the mind. We suppose that it is a representational system fed by perception, and organized internally somewhat like a filing cabinet. It contains ideas that stand for elements in the external environment; these ideas are connected in a complex interior web that reflects relations among the exterior elements they represent.³ Collectively, these files and their associated ideas embody the sum total of our explicit knowledge about the world. When we add to this image the suggestion that the mind employs the same trick for representing its contents that the filing cabinet in our example above employed, the result is a view of the mind in which every mental particular – every thought, experience, idea, and feeling, all the varied inhabitants of the mental realm – represents itself, lending those that already have an outwardly directed intentional significance a duality of content.

Regimentation

There's nothing inherently mysterious about dual content. We know how it works in espionage; agent A writes a letter with a mundane English content, but the letter contains a hidden message when interpreted according to code. This goes not just for linguistic media, but also for graphs, diagrams, sculptures, and models of all kinds. A graph may represent fluctuations of the stock market under one interpretation and present a silhouette of the Tucson Mountains under another. A child's stick and ball construction might represent part of his family tree under one interpretation and a hydrogen molecule under another. Whenever we have dual interpretations, we can ask questions about the semantic relations between them. There may be nothing of particular interest to say, as in the child's stick and ball construction, but if they overlap in subject matter, the text and subtext of a coded message may interact in interesting ways.

Let's call F under the mapping that associates files with their intentional objects, L1, and F under the mapping brought into play by the note in the F-file, L2. If we draw up a list of files in F, listing each beside its L1 and L2 interpretation, the semantic relations between them are easy to see.

² For convenience, we'll suppose that files have no semantically significant structure.

³ Perry and others have developed the metaphor into a proper theory. See, for example, *Reference and Reflexivity*, and *Knowledge, Possibility, and Consciousness* where he writes: "The basic picture is simple. We learn about objects through perceptions. When we perceive things, we have ideas of them (notions), that we associate with ideas for the properties we perceive them to have, creating a sort of internal file (a notion associated with ideas)...we retain these files when we are no longer perceiving the objects."

F-files	L ₁ -interpretation	L ₂ -interpretation
f ₁	a	q
f ₂	b	s
f ₃	c	a

The representational domain of L1 is represented by column 1; the representational domain of L2 is represented by column 2, and the first thing to notice is that there is overlap. Indeed, every item that appears in column 2 appears also in column 1, though not in general in the same row.⁴ So, for example, in this table, f1 under its L1 interpretation is coextensive with f3 under its L2 interpretation. Since the L2 interpretation of f3 is, of course, f3 itself, this means that f1, under its L1 interpretation, represents f3. A complete catalogue of these points of extensional contact would associate a self-description with each file in F, and would effectively interpret L2 in L1. It would locate F in the representational scheme of L1, the representational scheme it applies to the room as a whole. The form that the self-descriptions take will depend on what that representational scheme is. But if it is just the representational scheme of ordinary everyday thought, and if f1 is, for example, a 3/4 inch thick, red file in F's top drawer that stands for the table in the center of the room, a self-description would associate f1 with the "f1" file, the thickness of f1 with the "3/4 inch" file, the color of f1 with the "red" file, and so on. The fact that f1 is 3/4 inches thick, red, and located in F's top drawer would be represented by the pattern of connections between those files.⁵

Transformations between incommensurate representational schemes

We don't need to suppose that L2 or L1 are language-like. L2 clearly has more in common with media like maps or photographs. But we can still compare their representational domains, and whenever we have representations with overlapping domains, there will always be points of extensional contact.⁶ There are postcards that one can buy that place a map of Sydney beside photographs of parts of the city. Someone who knows the city can map the photographs into the map. If the resolution of the map were the same as that of the photographs, we could establish point-by-point mappings that would interpret each of the pictures in the map. We can establish similar points of extensional

⁴ Obviously, there are lots and lots of things that appear in column 1 that don't appear in column 2.

⁵ Files in ordinary cabinets have semantic structure. since we're supposing files are representational atoms, we externalize structure that is ordinarily present inside the files, so that instead of adding 'is red' to the A file to represent the fact that A is red, we draw a line representing the 'has' relation between the red file and the A-file.

⁶ Representational domains of languages are just universes of discourse. In general, we say that A falls within the representational domain of a medium M if it is represented by some semantically significant component of M.

contact between a verbal description of the perpetrator of a crime and a police artist's sketch, a portrait of the Smith family and a diagrammatic depiction of the Smith family tree, or a performance of Beethoven's *Moonlight Sonata* and a rendering in musical notation.

Start with a pair of schematic representations of F:

Diagram 1 (to be inserted)

Letting the left one stand for L2 and the right one stand for L1, draw arrows from the left into extensional equivalents on right.

Diagram 2 (to be inserted)

Now superimpose the two images, so that you have a set of arrows going from F into itself, mapping F's contents into files that represent them.

Diagram 3 (to be inserted)

What we've got here is effectively two separate, and self-contained, representations, embodied in the same structure, and a function that associates points of extensional contact. These points of extensional contact, once identified, create semantic bridges between the two 'levels' of representation. What makes cases in which we have one representation superimposed on another special is that in these cases, the mapping that associates extensional equivalents across representations maps the dually interpreted structure onto *itself*.

It might be convenient to have a name for the representational domain associated with the L2 content of a cabinet. I'll call this its 'presentational domain'. Every cabinet, and indeed any representational system of any sort, constitutes its own presentational domain. Closed systems have disjoint presentational domains. If F's universe were littered with cabinets all structured internally like F, each would have its own presentational domain and even if they were perfect internal duplicates the self-describing function that interprets L2 in L1 would vary from one to the next. The self-interpreting function, which locates each in the common representational scheme of L1 will be different because they are mapping distinct domains (distinct selves, in the reflexive sense) into L1.

Diagram 4 (to be inserted)

Transposing the picture

The picture is transposed as follows: we suppose that minds are closed representational systems, each of which constitutes its own presentational domain. We assume that some of its contents (perhaps all) carry both L1 (intentional) and L2 (presentational) contents.⁷⁷ Where we have distinct interpretations – or, we might say, ‘levels’ of representation – with overlapping domains, there are points of extensional contact, and these are given by level-bridging identities that relate *presentations* of mental particulars, with all of the substantiveness and determinacy that is characteristic of presentation (e.g., a shooting pain felt as a toe is stubbed), with intentional representations of them (the concept of a c-fibre firing in a brain located in space), characteristically lacking the substance and determinacy.⁸⁸

I have been describing these points of extensional contact in a metalanguage that permits reference to both presentational and intentional contents, but the question arises of whether there is a way for the system itself to represent them. The answer is yes. There are two tricks, both of which involve extending L1 beyond the resources of a first order medium. The first is quotation; the second is reflexive identification. The quotational model has been recently developed by Papineau, and there is precedent in the work of others. On the quotational model, one extends L1 in a way that allows it to mention expressions of L2 by producing a copy of the L2 item to be interpreted inside a special syntactic marker, and uses L1 to identify its extension.⁹⁹ So, for example, one imaginatively produces a phenomenal duplicate of a tactual/visual/auditory sensation, P. One places it in mental quotation marks to suspend its ordinary intentional significance and lets the expanded expression act as a name for the one it encloses. The quoted sensation is a syntactic, but not semantic, constituent of the expanded expression. One then forms the thought “P is what it is like for _____”, filling the blank with an objective description of goings on in the brain of JI, thereby bringing the presentational and representational schemes into partial extensional alignment, in just the way an English speaker does when she forms the thought “‘Cannelle’ refers to cinnamon”. The reflexive model is different. On the reflexive model, instead of producing a copy of the item to be interpreted *inside* the vehicle that carries L1 content, one ‘points at’ the thing itself. There are differences between the two techniques, and neither is eliminable. Both have uses that the other can’t perform. Where quotation allows you to mention a representation without using it, reflexive identification allows you to say

⁷⁷ I remain neutral on the scope of L1 and L2 content, i.e., on the question of which mental particulars have L1 content, which have L2 content, or whether all have both. This is a point of dispute that has been brought into the foreground of discussion lately by cognitive phenomenologists, like Horgan and Tienson, who have been defending the view that all phenomenal states have cognitive content, and *vice versa*.

⁸⁸ Levine 2001 introduced the terms substance and determinacy to express the difference between physical and phenomenal concepts. I am invoking his notions.

⁹⁹ See Papineau 2002.

something about items that are *in use*. No matter how much you put inside quotation, the quotation marks always fall outside the quoted content, but reflexive representation allows you to form a representation that can capture itself in its representational scope, without remainder.

L1-L2 equivalences

Most of us don't have very precise ideas about the goings on in our brains, and we interpret our phenomenal histories with vague, extrinsic self-locating thoughts of the form "this is the kind of event in JP's brain that is ordinarily caused by visual interaction with red things", "this is the kind of event that normally indicates trauma to the left knee", and so on. The vagueness of 'ordinarily' doesn't *ordinarily* matter, and these kinds of extrinsic descriptions are more useful for navigational purposes than intrinsic descriptions of the brain events they pick out (when you learn to drive a car, it's much better to know its parts by extrinsic functional descriptions like 'the button that causes such and such' or 'the wheel that controls so and so' than "knob of metal alloy, with radius *r*, in state *s*"). The extensional relationships are clearer if we imagine a neuroscientist who creates more direct bridging links between the physical map of his brain and self-presenting internal states by having lab assistants cause electrical stimulation in areas of his brain. He hooks himself up to the machine, tells them to induce various brain states and has self-locating thoughts like "*this* is a c-fibre firing, *that's* another c-fibre-firing, this is a d-fibre firing, this one now is a g-fibre firing," gradually charting the phenomenal similarity relations between various classes of physical events.¹⁰ Once he's got the bridging links in place, he reads his brain states off of his phenomenal states as easily and fluidly as someone who is adept with a map reads his objective coordinates off of the local landmarks. When he's in pain, he thinks to himself "there go my c-fibres", he describes tastes in terms of the regions of the brain they stimulate, he knows how to produce the pleasant stimulation of f-fibres, and so on. What the level-bridging identities do – both the loose, extrinsic level-bridging identities we employ, and the more precise counterparts of the neuroscientist – is bring the two representational schemes into alignment with one another.

Cognitive gaps

Now let's see what we can do to reproduce the cognitive phenomena. There are two, interrelated features of conscious experience that are invoked in every discussion of dualism as intractable obstacles to dualism.; the existence of cognitive gaps between physical and phenomenal properties that aren't resolvable by first order physical knowledge and the subjectivity of phenomenal concepts.

Whenever we have different ways of representing the same thing, we have potential cognitive gaps. When you have redundancy within a medium – e.g., coreferring terms within a language – cognitive

¹⁰ If physical and phenomenal concepts were drawn from incommensurable schemes, c-fibre firings in my brain would form a phenomenally motley crew of events. Phenomenal descriptions of physical similarity classes would be no more compressible than a list of their members.

gaps are resolvable by removing first-order ignorance about the domain.¹¹ A complete compendium of first-order truths about the world, expressed in English for example, will entail that water is H₂O, that the Evening Star is the Morning Star, and that the first Postmaster General of the United States is the inventor of bifocals.¹² And in general complete first-order knowledge of any domain expressed in M will yield all true extensional identities among terms of M. When you've got two separate media, the story is different. Each will be closed under the content-preserving transformation of its own states, but there will not in general be any way of getting from a term in one to a term in the other by content-preserving transformations defined over either. And this means that cognitive gaps between coextensive representations drawn from different media – e.g., between a description and a photograph, or between a geometric figure and an algebraic expression – first order knowledge expressed in either will not resolve cognitive gaps. Nothing internal to a sketch artist's rendering of the perpetrator of a crime, for example, will link the bulbous shape at the center of his portrait with the witness's term 'nose'. Nothing in an audio recording of Beethoven's *Moonlight Sonata* is going to link the first note with the blotch on the page in the written score. Nothing in your English doctor's knowledge of the human body will tell him the Hindu word for 'appendix'. One way of putting this is to say that extensional identities between terms within a medium supervene on first order truths, expressed in that medium; extensional identities between terms drawn from different media do not. And for this reason, nothing in the L1 contents of F-files, though they may include all the first order facts about the small universe that F inhabits, will entail a single level-bridging identity. What this tells us is that the fact that the cognitive and epistemic gaps between physical and phenomenal concepts are not resolvable by removing first order ignorance tells us nothing about their extensional relations. Physical and phenomenal concepts are concepts drawn from separate media and only metalinguistic knowledge will remove them.

The media in question, moreover, employ different representational schemes. By representational scheme, I mean simply a partition induced by the descriptive resources of a medium. Representational schemes are descriptive analogues of coordinate systems, they are sets of parameters that we use to classify objects by type. Sometimes a one-one mapping can be given between the cells of two representational schemes, e.g., between the representational schemes of English and Pig Latin. In that case, the schemes can be treated as notational variants, akin to coordinate systems related by a spatial translation. When schemes divide their domains into types that crosscut one another, there is a spectrum of degrees of commensurateness corresponding to the complexity of the function that transforms one into the other.¹³ A simple rule will transform English into Pig Latin but

¹¹ By 'first order', I mean non-meta-representational. First order truths are truths that don't explicitly mention ways of representing the subject matter. One can always resolve a cognitive gap between "A" and "B" by including "is called "B"" in the list of truths about A.

¹² A slightly more contentious, but illuminating case: complete first-order knowledge of arithmetic will entail that $2+2$ is 70-68, and $439+2$ is 441, but not that 2 is two.

¹³ Think of the transforming function as a translation manual. With the right specifications of language and so on, its complexity can be measured

something more complex is needed to translate Hebrew into Hungarian, a verbal description into a visual image, or an algebraic representation of figure into a visual shape.¹⁴ If no computable function can perform such a transformation, we say the schemes are *incommensurable*.

There are unresolved difficulties in understanding in general terms how to speak about extensionally coincident properties drawn from incommensurate schemes, difficulties that have troubled philosophers of science since Kuhn's remarks about incommensurability in the *Structure of Scientific Revolutions*. Properties are nodes in a network of logically related terms and many have held that their identity is partly dependent on their relations to other elements in the network from which they are drawn. If this is correct, it means that we can't properly talk about identity across networks even where we have extensional equivalence. The difficulties are compounded when the representational schemes are defined over overlapping, but non-coincident domains. What this brings out is that perhaps we shouldn't be talking about property identities at all. It's the extensional relationships that matter. Physicalism requires only that phenomenal properties fall within the scope of physical description, i.e., that the extensions of phenomenal concepts are, or are part of, the extensions of physical ones.

So, if we accept the dual content architecture, we have two levels of mental representation with overlapping domains, one superimposed on the other, and level-bridging identities, or equivalences that identify points of extensional contact. The level-bridging nature of these equivalences, and the fact that each level is internally closed, explains the persistence of cognitive and epistemic gaps in the face of complete knowledge of first order facts.¹⁵ But what of the renowned subjectivity of phenomenal properties?

Subjectivity

Although the problem of explaining qualitative character has been the principal focus of the philosophical literature on phenomenal consciousness, a few have maintained that the underlying problem lies with understanding subjectivity. Levine, for example, writes:

“what makes the problem of consciousness so hard is that we apprehend experience from a subjective point of view, and what is so apprehended cannot be simultaneously apprehended from the (more) objective point of view of physical theory ... it's not clear how subjectivity,

by the length of the shortest manual of translation.

¹⁴ When the representational domain is infinite, although there will be extensional overlap, there may not *be* a transforming function.

¹⁵ We could build completeness and internal closure into the definition of a representational medium if we adopt the following terminological stipulations. A representational medium is a set of information-bearing states together with a set of content-preserving transformations. Each medium is closed under the content preserving transformations of its states and equivalent states are connected by some chain of content-preserving transformations. The completeness condition, so stated, is too strong, but we might be able to find some suitably weakened version that captures the intuition that the terms of a medium should be semantically integrated with one another.

the cognitive relation constitutive of a point of view, can be explained in terms of [objective] properties.”¹⁶

He is following Nagel here, who has long maintained that the problem of absorbing phenomenal properties into physics is one aspect the problem of absorbing the first-person perspective into an impersonal view of the world. Phenomenal properties are, in some sense that has never been clearly understood, relational. There is something it is like *for you* when such and such occurs and something it is like *for me*, but nothing that any objectively described event is like *simpliciter*. Again, in Levine’s words:

“Not only is [seeing a red diskette case] a matter of some state (my experience) having some feature (being reddish) but, being an experience, its being reddish is “for me,” a way it’s like for me, in a way that being red is like nothing for – in fact is not in any way “for” – my diskette case.”¹⁷

In the closing paragraphs of his book, it is the subjectivity of conscious experience that Levine settles on as the real outstanding difficulty for the physicalist

“the source of the problem with providing a physical explanation of qualitative character resides in the subjectivity of conscious experience.”¹⁸

Here is what subjectivity amounts to and how it is explained on the dual content model. Subjects, like filing cabinets, are closed representational systems. To each, there corresponds a presentational domain, i.e., a collection of properties that are exemplified by its representational states. These properties are the referents of ‘this’ in level-bridging identities. When our neuroscientist hooked up to his brain stimulator says ‘*this* is what it like when c-fibre firing occurs (in my head)’, he maps ‘c-fibre firing’ onto a property exemplified in his presentational domain. Each of us does the same. I identify points of extensional contact between physical properties and properties exemplified in my presentational domain. Y identify points of extensional contact between physical properties and properties exemplified in your presentational domain. We use the public language to talk to one another.¹⁹ These internally exemplified properties will serve for each of us as points of reference in understanding descriptive vocabulary. When I am told that something is red, sour, itchy or hot, for example, I relate it to properties exemplified in my presentational domain, exemplars drawn from past experience and connected to ‘red’, ‘sour’, ‘itchy’ and ‘hot’ in ‘*ah, so that’s*

¹⁶ p. 177.

¹⁷ *Purple Haze*, p. 6-7.

¹⁸ p. 8-9.

¹⁹ Our filing cabinets, of course, can do the same. F will establish points of extensional contact between the common language and the properties exemplified by the paper and ink that it contains. F* will establish points of extensional contact between the public language and properties exemplified by the paper and ink on which it is written. F will write “this is blue”, ‘here is a triangle’, F* will write ‘this is black’, ‘there is a circle’. Each describes its own contents in the public language. There may or may not be overlap in the properties each exemplifies.

what redness/sourness/itchiness and heat are like-thoughts, or, more explicitly, 'ah, so that's what 'red', 'sour', 'itchy' and 'hot' refer to'.

Now let's add to this a consequence of the causal theory of perception fact, *viz.*, that the only way we have of characterizing properties exemplified inside presentational domains using the shared language is by causal relations to properties in the public environment.²⁰ The reason that is so has to do with the way the public vocabulary is interpreted.²¹ It will follow that when it comes to describing internally exemplified properties the relativization to a presentational domain is ineliminable. In fact, it will follow there is an implicit frame dependence in all descriptive vocabulary that only becomes explicit when it is self-applied. It makes no sense in general to say that an event is *like* this or that; one has to say 'like this or that *for whom*'. If it's an event that falls outside one's presentational domain, one says what kind of causal impact it makes on properties that are exemplified therein. If it's an event that falls inside one's presentational domain, one asserts a level bridging identity, 'this is a c-fibre-firing', 'there go my y-fibres again'...or some such thing.

more cognitive phenomena

There are other noted peculiarities of the mind's representation of its own contents that are also reproduced by the dual content architecture.

1. the cognitive immediacy that characterizes awareness of one's own occurrent conscious mental states contrasts with the cognitive distance we feel from physical properties; its own occurrent states are present to the mind in a manner that merely intentional objects are not.²²

2. the fact that one cannot coherently entertain doubt about the existence or character of such states; one cannot coherently entertain doubt about the existence or character of such states.²³

3. the fact that one cannot doubt that presented states are *one's own*: "I wonder if this thought is mine or someone else's" is an incoherent thought."²⁴

²⁰ I won't provide any defense of this claim here, but see SS. It's a familiar consequence of the causal theory of perception.

²¹ There is no analogue of this for our filing cabinets, because we have not built in any analogue of perception or said anything about how the public vocabulary is interpreted. For discussion, see *The Situated Self*.

²² And that makes it impossible to seriously entertain doubts about their existence or their intrinsic character, of a sort that are possible with respect to merely intentional objects.

²³ The question 'can I fail to contain a representation of something I contain, or can I misrepresent myself as containing something I do not in fact contain?' for F is the question 'can I fail to contain something that I contain, or can I contain something I fail to contain?'. To which the answer is always 'no'. Still, however, F can misrepresent the contents of other files, and they, in their turn, can misrepresent F's contents.

²⁴ It is important to distinguish the impossibility of reflexive from intentional misrepresentation of one's own thoughts. One can mistakenly intentionally self-ascribe someone else's thought, as the deluded Heimson,

4. the relational character of all of these features, i.e., the fact that no mind has similar guarantees with respect to the contents of minds other than itself.²⁵

All of these arise directly from the special character of reflexive representation. The possibility of ignorance and error arises in the gap between concept and property, i.e., between what is represented and what is *doing* the representing. Reflexive representation closes that gap.

Closing

I have argued that epistemic and cognitive gaps on which influential arguments for dualism are premised are artifacts of the reflexive nature of identification of phenomenal properties. They don't support dualism, and they don't show anything, one way or the other, about the completeness of physics. But, for the record, as reductive accounts of what it *is* to be conscious self-representational analyses seem to me as misguided as dualism. When the dust clears, the central mystery of consciousness remains untouched – its undefinability and incommunicability - not because consciousness is a special, non-physical quantity or substance, but because the degenerate character of reflexive knowledge leaves no room for a positive understanding what consciousness is. This needs to be spelled out carefully, but perhaps on another occasion.

for example, presumably did when he self-attributed the thoughts that originated in Hume's cranium.

²⁵ It even provides us with an explanation of Moore's Paradox, viz., the fact that, although it is certainly possible that p is the case, though one doesn't believe that p, one cannot consistently form the thought "p and I don't believe that p". The presence of p in one's belief box *represents* the presence of p in one's belief box, so one cannot have that belief without simultaneously representing oneself *as* having it.