

SOME SEMIOTIC CONSTRAINTS ON METAREPRESENTATIONAL ACCOUNTS OF CONSCIOUSNESS

55

Marc Champagne

York University (Toronto)

“Representation” is one of those Janus-faced terms that seems blatantly obvious when used in a casual or pre-theoretic manner, but which reveals itself far more slippery when attentively studied. Any allusion to “metarepresentation”, it would then seem, only compounds these difficulties. Fortunately, there is a rich discursive strand which makes it its chief business to carefully unpack all that is implied by the notion of something representing another. Pursuant with this *semiotic* vocation, we propose to investigate the metarepresentationalist account which has been invoked by philosophers and cognitive scientists as a promising way to explain the nature of human consciousness.

Metarepresentational accounts of consciousness typically fall into two basic families. On the one hand, we find theories that center on “higher-order *thoughts*” or “HOTs”; while others employ the notion of “higher-order *perceptions*” or “HOPs”. However, we shall largely disregard this divergence, and will direct our efforts on the superordinate class involving the broader notion of “higher-order *representation*”, here taken to encompass thoughts, perceptions, or whatever. Indeed, despite their differences (which, naturally, strike their respective proponents as paramount), both HOTs and HOPs mobilize a common insight: conscious states involve a folding of the mind onto itself, as it were. This can be read as a substantive claim about the nature of consciousness: where there is no higher-order folding, there is no consciousness in the demanding sense.

We shall offer neither a defense nor a critique of this hypothesis, much less an assessment of the feasibility or challenges that face inquiries into consciousness generally. Taking the metarepresentationalist framework in its roughest outline as our point of departure, we shall instead try to articulate in an informative manner some of key “structural” features that appear binding for any such theory. This, then, will not be a survey of the literature on the topic but rather a tangible contribution to that body. The overarching hope is that by bringing a properly semiotic reflection to bear on the seemingly innocent notion of metarepresentation, we shall be making explicit some of the constitutive conditions that perforce constrain the elaboration of such frameworks, thereby flagging potentially otiose avenues of research.

First constraint

The re-application involved in metarepresentationalism is not prejudiced in favor of any particular building-block. To be sure, the theorist wishing to produce a complete account must eventually settle on the candidate material she deems most suitable for duplication. But metarepresentationalism *per se* makes its entry onto the scene only once a representational relation is directed not at the world but at yet another representational relation. Let us then provisionally define a representation as “something which stands for something else”—to adopt the succinct medieval formula “*aliquid stat pro aliquo*”. Accordingly, we can describe a *higher-order* representation as something that stands for something else which stands for something else.

One of the crucial features which must not be corrupted on pain of there no longer being any rationale for talk of “higher” levels is the kinship both representations share. The point is a general one, and can be variously rendered. If, say, the French language is described in an English manual, it makes little sense to say that the manual in question is somehow of a “higher” order than the foreign tongue it deals with. The idiom of first- and higher-orders would be warranted, however, in the case of a French manual on the French language. This, intuitively, is what is meant by kinship.

We can follow Klinkenberg (1996: 57) and speak of the relata as being “homosemiotic” (as opposed to “heterosemiotic”). Lacking a homosemiotic kinship, we simply have no sufficient grounds to start elevating our theoretic depiction into “meta” levels. Invoking parsimony, we should rather say that an English manual on French is simply a “something” which deals with “something *else*”—even though the French language clearly fulfills a representational duty of its own. In such a heterosemiotic case, we would have something that stands for something else (that in turn stands for something else), but the relations would be strictly horizontal, so to speak; and it would only be superfluous to think the transitivity at hand in terms of ascending levels. In other words, a “heterosemiotic metarepresentation” is a contradiction in terms (a point ostensibly missed by Klinkenberg). This precept rests on the fact that, in the heterosemiotic case of English-to-French, incorporation of first- and higher-orders would be redundant: the very difference between the representations robustly underwrites their discernment.

The matter is altogether different, however, when a homosemiotic kinship does run across the parties. An English manual dealing with the English language can (and in an important sense *must*) be understood in terms of “meta”-levels—and this for the very same reason that different languages needn’t (and mustn’t). It is thus essential that (1) *whatever mechanism or cognitive process a metarepresentational account favors not differ in any fundamental way once it is redirected onto itself*. The exemplar of this first constraint is John Locke, who availed himself solely of empiricist resources and construed introspection as a perception directed at extrospection (that is, at a perception directed at the world). For all its simplicity, this sort of re-employment is a powerful explana-

tory fulcrum, as it suggests that one could run an exhaustive account of consciousness on a single mechanism, thereby allowing us to exploit processes that are better understood.

Second constraint

Suppose, then, that a metarepresentational theory elects to build its account of consciousness from fairly sedate first-order materials. Does this mean that the constraints imposed on its theorizing are *eo ipso* limited simply to those that come with its chosen building-blocks? As we shall endeavor to demonstrate, *further* constraints ensue when the said materials (e.g., thought or perception) are folded onto themselves to obtain a multi-layered architecture. Specifically, the transitivity from one representation to another is drastically affected depending on how one construes the idea of something being re-presented. In order to illustrate what’s at stake, let us borrow a very suggestive passage from Alex Byrne (1997: 103):

Consciousness is the subject of many metaphors, and one of the most hardy perennials compares consciousness to a spotlight, illuminating certain mental goings-on, while leaving others to do their work in the dark. One way of elaborating the spotlight metaphor is this: mental events are loaded on to one end of a conveyer belt by the senses, and move with the belt—perhaps changing as they go—towards a fixed circle of light, which does not completely cover the width of the belt. Some mental goings-on fail to pass through the illumination, in which case they never become conscious. But others are illuminated, and thereby enter one’s consciousness. Beyond the spotlight, at the other end of the conveyer belt, lies the filing cabinet of memory, into which some of the more garish or lurid of the belt’s contents fall.

There are several ways to construe the various components which give saliency to a store of mental contents, and the merit of the metaphor tendered by Byrne is of course open to debate. There is the thorny issue, for example, of whether it even makes sense to countenance “unlit” mental contents and/or activities. Although one could shave off those excess portions of the conveyer belt not enjoying cognitive attention, we shall conveniently spare ourselves this task by limiting our inquiry to that subset of contents that *are* represented, staying mum on what might or might not lurk beyond such a scope. In keeping with this disregard of dispositions which (for one reason or another) fail to garner due cognitive attention, we proceed from a baseline of fully-fledged conscious states. Such a state involves at least two parties: a (salient) content and a “spotlight” which represents it (hence the salience). Although we shall be called to provide a minimal positive account of what such a nondescript “spotlight” consists in, it will greatly disburden our exposition to leave the metaphorical stand-in unexplained for the time being.

The generic structure we want to explore, then, is that of a spotlight terminating on a content of some sort. Keeping in mind the “kinship” constraint outlined earlier, whatever metarepresentational structure we erect must not dif-

fer in nature from this—save of course in virtue of the hierarchy which helps to distinguish the different levels. All told, the goal is one of economy: we want to obtain complex patterns like consciousness by reiterating a less onerous elementary structure. Let us therefore ascend from a first-order level and reapply the basic representational structure, taking yet another spotlight and turning it onto the first spotlight-to-content couplet. Thus, if the initial representation was originally taken to be a thought, we obtain a HOP which effectively “thinks of a thought”. Similarly, construing the first-order representation as a perception means that to be a HOP is to “perceive a perception”. Given that the representation which was employed to augment the initial situation terminates on a representation of like nature, it makes perfect sense to label these “second-” and “first-order” representations, depending on the role.

The question we may now ask is: what is this higher-order representation a representation *of*? At first blush, the answer to this query seems plain: more of the same. Indeed, the very vocabulary of first- and higher-orders guarantees as much. We can therefore make the concern less trivial by recasting it as follows: *what element* of the first-order representation does the higher-order representation terminate on? In other words, assuming representations to be irreducibly complex relations, onto which “part” of a first-order representation does a higher-order representation terminate—the content, the “spotlight”, or both (i.e., the spotlight illuminating the content)?

Let us consider the first option. What would happen if a higher-order representation represented only the same (presumably worldly) content which a first-order representation represents? This would entail that both representations would be on par *qua* levels. Indeed, in such a situation, B would stand for A, and C would also stand for A; but since neither B nor C would stand for each other, there would be no warrant in describing one as being of a “higher” order than the other. This effectively ejects us from the sphere of metarepresentation altogether and brings us to *synonymy*. Quine’s misgivings notwithstanding, that’s not inherently a bad thing; but it does entail that (2) *a higher-order representation cannot represent a first-order representation’s content merely*, on pain of dropping down to that first-order level and becoming a mere synonym terminating on a common (worldly) content. This then, can be considered our second semiotic constraint.

Third constraint

If we agree that a higher-order representation cannot represent only that which a first-order representation represents, we are led to conclude that anything which is deservedly “of a higher-order” perforce represents either the first-order “spotlight” that illuminates a content or both. Given this conclusion, the question arises: what might such a “spotlight” be, that it may in turn be illuminated? Clearly, the time has come to put more flesh on the metaphor.

Proceeding cautiously, we seem compelled to concede this much: whatever represents a content at a first-order level must itself demonstrate enough

reality to be subsequently represented by a higher-order duplicate—that is, to be the terminus of another representation. No matter how queer it may seem, it is important to stress how crucial this minimal ontological commitment is to the success of a metarepresentational account. No one loves a homunculus, especially when it looks a lot like a Cartesian Ego. There is thus an understandable tendency to recoil from this and maintain that whatever represents a content in a first-order level admits of no substantiality whatsoever. According to this position, contents are apprehended by a “view from nowhere”—to borrow an evocative expression of Thomas Nagel’s and put it to a different use. This gloss would thus prevent a higher-order representation from falling back to mere synonymy by having it represent the first-order content and an attached illumination without tangible origin.

As far as we can tell, there is nothing inherently wrong with this sort of reply. The only problem is that the idea of an attentive saliency without provenance—a content needing no representing—is a luxury metarepresentational theories simply cannot afford. Indeed, just as it is logically necessary that a higher-order representation represent *more* than its first-order representation’s content (*pace* our second constraint regarding synonymy), so this “more” must be capable of being represented in turn. The cash-value of the “view from nowhere” view is that it allows contents to sparkle without there being any representing party doing the work. Contents would thus have their cognitive saliency intrinsically. That’s all fine, as far as it goes; but it gives the metarepresentationalist literally *nothing* to reapply.

If instead of a representational complex the content of each conscious state would be a unary monad wholly accounting for its own “representational” power, then there would be nothing for a “higher-order” to represent. Strictly speaking, we wouldn’t even have cause to describe such intransitive contents as being representations—“presentation” would be a far more apt description of such a situation (see for example Gamble 1997). Correspondingly, if contents were “viewed from nowhere”, a third-party observer (broadly construed) witnessing this strange sort of “viewing” would have nothing to see save the content at hand. Yet if this is so, then the best one can hope to achieve theory-wise is yet another presentation of the same content. As with synonymy, this forecloses the possibility of ascending to higher orders.

We routinely operate on a transparent level that rarely directs its contemplative gaze at the means by which the denizens of everyday life phenomenologically manifest themselves before us. In other words, we rarely take the *obiectum quo*—the “object by which”—and make it an *obiectum quod*—the “object that”. But if the above considerations are correct, metarepresentationalism simply cannot take this transparency at face-value. The illuminated contents Byrne alludes to must have their light bulbs, thought thoughts must have their thinkers, perceived perceptions must have their perceivers—and so on. Otherwise, the metarepresentationalist project simply has no traction. The third semiotic constraint a metarepresentational account of consciousness must negotiate, then, is that (3) *whatever does the representing on a first-order cannot be*

nothing, since this would mean that higher-order representations have nothing to represent. Thus, when we rigorously analyze the notion of metarepresentation, we see that it implies a commitment to the reality (however minimal) of *both* parties in the Latin formula “something stands for something else”.

Fourth Constraint

Let us recap the constraints we have extracted thus far. A metarepresentation must not differ in kind from whatever representation it rises above. Moreover, it cannot simply reprise the same (worldly) content as a first-order representation, on pain of being knocked off that meta-level. Indeed, for a representation to truly belong to a “higher order”, it cannot represent merely a first-order content but also the first-order *representation of* that content. This implies that contents cannot intransitively account for themselves. A theory meets this requirement if it countenances sign-vehicles of enough ontological substantiality for them to be represented in turn. So the *obiectum quo* must be an “*obiectum*” after all.

With these provisions in place, our final constraint follows directly. Indeed, if every representation comprises a sign-vehicle and a content which this vehicle points to—as “something” that stands for something other than itself—then (4) *there is no principled reason within the metarepresentational framework itself why these abstract roles could not be compounded over and over.* In other words, if every illuminated content has a light bulb which can in turn be illuminated by another, then the relation can be multiplied several times. We see, then, that the construal of representation as “something standing for something else” is not satisfactory for metarepresentational accounts, and needs to be augmented (following C. S. Peirce) to a triadic “something standing *for* something *to* something”. This third party is yet another thought, or another perception, depending. But in either case, contemporary philosophy of mind, in spite of its relative ignorance of semiotic theory, seems to have arrived at an outlook quite analogous to *semiosis*, where signs can beget further signs.

This is a remarkable convergence, and deserves to be noted. Yet while semiotics has embraced and even celebrated the potentially “unlimited” character of semiosis (most notably with Eco 1976), advocates of metarepresentationalism in philosophy have until now tended to limit their hierarchical structures to two or three higher-order levels at most. We fail to see what might motivate such qualms vis-à-vis the prospect of far longer semiotic series. Perhaps one could reply that a regress of metarepresentations will inevitably be halted by some inherent biological or psychological limitations. That’s indeed liable to be the case. Given the energy expenditure associated with such growth, we can rest assured that if the embodied agency which drives a metarepresentational architecture does not have any incentive to weave that semiotic fabric beyond its needs, it won’t.

But the objection from limitations is usually more crude, and seeks to exploit the fact that agents can fathom only so much—subitizing being restrict-

ed in most humans to the non-inferential apprehension of 7 ± 2 items at a sitting (i.e., the old “how many things were on the table without counting” bit, pace Miller 1956). However, invoking this tangible limitation in the context of metarepresentational theories is misguided, as it effectively makes the thinking ego into a supplementary party, an unaccounted-for searchlight viewing the chain of representations as a spectacle that would be devoid of meaning were it not for its synoptic gaze. Not only would such a line of reasoning lapse into a familiar quagmire (one exposed in the wake of the “multiple drafts” model put forth by Dennett 1991: 101–138), it would grossly miss the point of metarepresentationalism as a substantive account of what consciousness *is*.

If we are to realize Rosenthal’s hunch (1997: 747) that metarepresentation is “our best hope” for circumventing the usual fare of Cartesianism whilst reaping something close to our macroscopic sense of consciousness, surely more than a couple iterations are needed. But just as a materialist expectation of what it means to be “real” prevents metarepresentational theories from fully embracing the vehicles they are bound to countenance, so a tacit mechanistic worldview seems to prevent them from exploring those fruitful explanatory avenues that are opened-up when one ceases to artificially impose a ban on the generation of higher-order levels. However, no perpetual motion machine is turned on the moment a representation is reapplied onto another. Indeed, the whole point of semiosis as a truly *post-modern* idea (see Deely 2001) seems to be that it is not reducible to efficient causation in the narrow sense (though it can of course subsume this).

Sobriety is recommended in heeding this last insight. Specifically, we must guard against the *non sequitur* fallacy that equates any challenge to efficient causation’s exhaustiveness with an endorsement of final causation. Such a view holds that since we are not pushed from the beginning of time like dominoes, we are pulled inexorably towards an end of time. Neither of these grand schemes does justice to the facts at hand. The potentially open-ended character of consciousness implied by metarepresentationalism is thus no license for exploding that very concept and positing occult attractors outside the spatial and temporal purview of the individual agent possessing that consciousness. Of course, to the extent one purges teleology of any Hegelianism and assigns it a more humble scope, then this warning can be ignored. Still, in the context of the ongoing “biological turn” in semiotics (i.e., the influential and probably right-headed thesis that sign-use is the distinctive trait of living things), it is necessary to augment our fourth constraint and stress that *discarding a disenchanting perpetual motion machine only to embrace an enchanted one is no advance*, and would betoken a profound disservice to the field. For biology is not “the study of life”, as it is often said without ado, but the study of *lives*—and these are very much *finite* (in fact, that is what gives rise to “ends” in the first place). And since consciousness is always situated and embodied, it too shall perforce be finite, no matter how complex and stratified it gets. So it is safe to assume that a successful metarepresentational account of consciousness will involve more than one iteration, but less than an infinity.

References

- BYRNE, Alex.
1997. "Some Like it HOT: Consciousness and Higher-Order Thoughts", *Philosophical Studies* 86, 103–129.
- DEELY, John N.
2001. *Four Ages of Understanding* (Toronto, Canada: University of Toronto Press).
- DENNETT, Daniel C.
1991. *Consciousness Explained* (Boston, MA: Little, Brown and Company).
- ECO, Umberto.
1976. *A Theory of Semiotics* (Bloomington, IN: Indiana University Press).
- GAMBLE, Denise.
1997. "P-Consciousness Presentation / A-Consciousness Representation", *Behavioral and Brain Sciences* 20.1, 149–150.
- KLINKENBERG, Jean-Marie.
1996. *Précis de sémiotique générale* (Brussels: De Boeck-Wesmael).
- MILLER, George A.
1956. "The Magical Number 7, Plus or Minus Two: Some Limits on our Capacity for Processing Information", *Psychological Review* 63.2, 81–97.
- ROSENTHAL, David M.
1997. "A Theory of Consciousness", in *The Nature of Consciousness: Philosophical Debates*, ed. Ned Block, Owen Flanagan, and Güven Güzeldere (Cambridge, MA, and London: MIT Press), 729–753; adapted from Center for Interdisciplinary Research (ZiF) Technical Report, no. 40 (University of Bielefeld, Germany, 1990).