The Problem of the Rock
and the Grammar of Consciousness

Lajos Brons

Nihon University and Lakeland University

Abstract

The "Problem of the Rock" (PoR) is a famous objection to Higher-Order (HO) theories of consciousness. According to PoR, the HO theorists' claim that a mental state is conscious iff there is a higher-order mental state about it implies that a rock is also conscious iff there is a higher-order mental state about it. In this paper I show that this argument confuses two grammatically distinct attributions of consciousness, and that if the consequent equivocation fallacy is avoided, PoR is either a straw man argument or has an unproblematic conclusion.

According to so-called Higher-Order (HO) theories of consciousness, a mental state is conscious if and only if there is a suitable higher-order mental state (either a thought or a perception) about it. One of the best known objections to HO theories was first put forward by Alvin Goldman over two decades ago:

> A rock does not become conscious when someone has a belief about it. Why should a first-order psychological state become conscious simply by having a belief about it? (Goldman 1993: 366)

Leopold Stubenberg (1998) called this 'the problem of the rock' (PoR). Robert van Gulick (2000) used the term 'generality problem' for a nearly identical objection, differing mainly in its substitution of a desk lamp for Goldman's rock.

HO theorists have responded to PoR in a number of different ways. (See Gennaro 2005 for an overview.) The most obvious response is to deny the validity of the generalization from mental states

to rocks (or desk lamps). In this paper I will argue that there is a more fundamental problem for PoR, however: it depends on an equivocation of two grammatically distinct attributions of consciousness.

Before we can properly assess PoR, there are two minor defects in Goldman's original phrasing of the problem that need to be corrected. Firstly, as Rocco Gennaro (2005) has pointed out before, the term 'belief' is somewhat inappropriate here. HO theories claim that the higher-order state is either a thought (in Higher-Order Thought theories) or a perception (in Higher-Order Perception theories), but not that it is a belief. (On the other hand, if in some HO theory thoughts and/or perceptions are considered to be kinds of beliefs, then for that theory the term 'belief' might be acceptable.)

Secondly, it is not entirely clear how 'becoming conscious' should be understood here (i.e. temporally, causally, etc.). The most neutral reading equates it with 'being conscious', and that is probably the most charitable reading as well.

Taking these corrections into account, the second sentence of the above quote should be understood as an interrogative version of the claim that:

(1) A mental state *x* is conscious if and only if someone has a thought about/perception of *x*,

which summarizes Goldman's apparent interpretation of HO theories. Whether this interpretation is entirely correct is debatable. At the very least, the word 'suitable' needs to be inserted before 'thought'. Nevertheless, in order to focus on the aforementioned more fundamental problem, I will let it pass and assume that (1) is an acceptable summary of the main claim of HO theories.

Goldman's point is that, lacking a good argument to the contrary, we should be able to generalize (1) to:

(2) Any *x* is conscious if and only if someone has a thought about/perception of *x*,

from which it follows that:

(3) A rock is conscious if and only if someone has a thought about/perception of that rock.

However, (3) is false because rocks are not conscious, and therefore,

(1) is false. Or at least, that is—supposedly—what the enthymematic conclusion of PoR should be.

In one of the earliest responses to PoR, David Rosenthal (1997) appealed to a distinction between transitive consciousness of something and intransitive being conscious, and charged PoR of being "just a disguised version of the doctrine that being intransitively conscious is an intrinsic property" (1997: 739). Intrinsic properties are non-relational—that is, they do not consist (even partly) in "bearing a relation to something else" (1997: 736). If PoR would indeed presume that intransitive consciousness is intrinsic, then this would be begging the question against HO theories, because those assume consciousness of any kind to be a relational property. However, it has been disputed that PoR depends on this assumption. Alex Byrne (1997), for example, argues that it is more likely to follow from PoR than lead to it. (See also Gennaro 2005.) The distinction between transitive and intransitive consciousness, on the other hand, is essential to understanding PoR, but insufficiently fine-grained.

The transitive/intransitive distinction is (at least originally) a grammatical distinction between two kinds of verbs. Transitive verbs take two arguments. For example, 'John is eating a steak' and 'Jane opens the door'. Intransitive verbs take only one argument: 'John is eating' and 'the door opens'. The verbs in these examples, 'eating' and 'opening', are really ambitransitive verbs, meaning that they can be both transitive and intransitive. If we bracket 'of' in 'being conscious (of)', then that phrase appears to be similarly ambitransitive.

There are (at least) two kinds of ambitransitive verbs, which is illustrated in the two examples in the previous paragraph. The agent of intransitive 'eating' is the agent of transitive 'eating', while the agent of intransitive 'opening' is the patient of transitive 'opening'. For that reason, 'eating' is called an 'agentive ambitransitive' and 'opening' a 'patientive ambitransitive'.

Usually, the intransitive form of an ambitransitive verb can have either the agent or the patient of the transitive form as its sole argument, and is thus either an agentive or a patientive ambitransitive, but there are also a few verbs that are doubly ambiguous. Such verbs are ambitransitive, but unlike most ambitransitives, they can take either argument as the sole argument of the intransitive form. For example, transitive 'Linda is flying this plane over Siberia' has two

related intransitive sentences: 'Linda is flying over Siberia' and 'this plane is flying over Siberia'. Such verbs appear to be rare, however, but 'being conscious (of)' works (more or less) like such a verb, and this is a source of (potential) confusion. As an illustration, consider the following three sentences:

(a)   Jake is conscious of his thought about breakfast.

(b)   Jake is conscious.

(c)   The [i.e. Jake's] thought about breakfast is conscious.

In (a) 'being conscious (of)' is transitive; in (b) and (c) it is intransitive, but while it takes the agent of the transitive form as its argument in (b), the argument of (c) is the patient of the transitive form. In other words, (b) attributes (some kind of) consciousness to an agent, and (c) to a patient. For that reason, I will refer to these two different kinds of intransitive consciousness as agent-consciousness and patient-consciousness, respectively. There is a further complication, however.

The transitivity of 'being conscious of' in (a) can be represented symbolically as $\langle 1,1 \rangle$ in which the order is $\langle \text{agent,patient} \rangle$ and in which '1' marks presence of an argument and '0' marks omission. In the same notation, the transitivity in (c)—that is, of patient-consciousness—is $\langle 0,1 \rangle$. In that sentence, the agent is omitted and the patient is promoted to grammatical subject similar to the passive voice. Analogously, the transitivity in (b) could be interpreted as $\langle 1,0 \rangle$ similar to the antipassive voice (which does not exist in English, but that does not matter here), but that is not the only possibility.

A genuinely ambitransitive verb is $\langle 1,1 \rangle$ in its transitive form and either $\langle 1,0 \rangle$ or $\langle 0,1 \rangle$ in its intransitive form (the first if its agentive, the second if it is patientive), but there also are quasi-ambitransitive verbs that act like ambitransitive verbs grammatically, but that have $\langle 1 \rangle$ intransitive forms that differ subtly in meaning from their transitive forms.

The difference between $\langle 1,0 \rangle$ and $\langle 1 \rangle$ can be illustrated by means of a comparison of 'eating' and 'falling'. If John is eating, then he is eating something, but if the ball falls, then the ball is not falling anything (and neither is anything falling the ball). In other words, the ambitransitive verb 'eating' implies a hidden argument even in its

intransitive form—that is what the '0' in ⟨1,0⟩ means. But a genuinely intransitive verb—like 'falling'—has no such implication, and should thus be represented as ⟨1⟩.

'Walking' and 'flying' are examples of quasi-ambitransitives. The first can be transitive as in 'Sophie is walking the dog' or 'Hanako is walking a long-distance trail' and intransitive as in 'Margaret is walking', but the last sentence does not imply that Margaret is walking something (i.e. a pet, or some route or path), and therefore, the transitivity of 'walking' in that sentence is ⟨1⟩. The same applies to 'the pilot flies the plane' and 'the bird flies'. Importantly, there are subtle differences in meaning between the transitive and intransitive forms of 'walking' and 'flying'—and there must be, given that the intransitive forms have no implied patients—and perhaps it can be argued for that reason these are not really grammatically different forms of the same verb but closely related homonyms. For the same reason, if agent-consciousness is ⟨1⟩ rather than ⟨1,0⟩, then it does not express the (exact) same concept as transitive consciousness, and is probably not reducible to it.

While transitive consciousness and patient-consciousness can be easily characterized as ⟨1,1⟩ and ⟨0,1⟩, respectively, it is not clear—and certainly not uncontroversial—whether agent-consciousness is ⟨1,0⟩ or ⟨1⟩.[1] Furthermore, different positions with regards to HO theories most likely also differ in their interpretation of agent-consciousness as either ⟨1⟩ or ⟨1,0⟩. And if that is the case, then there is no non-question-begging definition of agent-consciousness. Nevertheless, even if we cannot uncontroversially define agent-consciousness, there is at least one important feature that follows in either interpretation.

Patient-consciousness is ⟨0,1⟩ and is, therefore, defined as in (4). If agent-consciousness is ⟨1,0⟩, then it just omits the other argument and is, therefore, analogously defined as in (5*). (The asterisk marks that (5*) is true only if agent-consciousness is ⟨1,0⟩.)

(4) ∀x [x is patient-conscious ↔$_{def.}$ ∃y [y is conscious of x]]

---

[1] My own opinion on the matter is that the notion of ⟨1⟩ intransitive consciousness is confused nonsense (and the main cause of most philosophical problems about consciousness), and thus that any occurrence of intransitive consciousness is either ⟨1,0⟩ or ⟨0,1⟩, but I will not defend this opinion here.

(5*)∀x [x is ^(1,0)agent-conscious ↔_{def.} ∃y [x is conscious of y]]

In either case, intransitive consciousness is defined in terms of a primitive concept of transitive consciousness (i.e. 'being conscious of', 'being aware of', etc.). No such definition of agent-consciousness is possible if 'being conscious' in that sense is really ⟨1⟩ intransitive, because in that case, it does not express the (exact) same concept as the transitive variant. However, even in that case there appears to be an uncontroversial conceptual relation with transitive consciousness: it seems undeniable that being agent-conscious at least implies the possibility of being conscious of something:

(6) ∀x [x is agent-conscious → ◇ (∃y [x is conscious of y])]

If it is accepted that p implies ◇p, then (6) also follows from (5*), and consequently, (6) is true regardless of whether agent-consciousness is ⟨1⟩ or ⟨1,0⟩.

(6) obviously implies a restriction to what kinds of things can be agent-conscious, namely things that can be transitively conscious of something, but no similar restriction is implied by (4): anything can be patient-conscious, and something can even be agent- and patient-conscious at the same time. If two people are having a conversation, they are both simultaneously agent- and patient-conscious.[2]

The first premise of the Problem of the Rock (PoR) is a summary of HO theories, repeated here for convenience:

(1) A mental state x is conscious if and only if someone has a thought about/perception of x.

PoR is about intransitive consciousness, and therefore, 'is conscious' in (1) is either agent-consciousness or patient-consciousness. Because the subject of (1) is a mental state, 'is conscious' can be agent-consciousness only if mental states are a kind of thing that can be conscious of something (i.e. not just about something, but conscious of that thing). As far as I know, no HO theorist ever claimed that mental states can be conscious agents or can be conscious of things

---

[2] If Jane is conversing with John, then (a) Jane is conscious of John, and therefore (a1) Jane is agent-conscious and (a2) John is patient-conscious; and (b) John is conscious of Jane, and therefore, (b1) John is agent-conscious and (b2) Jane is patient-conscious.

(in this sense), and therefore, this interpretation would misrepresent HO theories. Furthermore, even if HO theories would claim that mental states can be conscious agents, it would not follow that anything can be a conscious agent as asserted by (2), because (6) implicitly restricts agent-consciousness to particular kinds of things. Consequently, if 'being conscious' in PoR is agent-consciousness, then the argument is not just a straw man (because it misrepresents HO theories), but invalid as well.

Alternatively, if 'is conscious' in (1) is patient-consciousness, then from (4) it follows that someone (or something) is conscious of that state, and that appears to be exactly what HO theorists are claiming. Hence, we can disambiguate (1) as follows:[3]

(7) A mental state x is patient-conscious if and only if someone has a thought about/perception of x.

From (4) it follows that the left-hand side of the biconditional in (7) implies that there is some y that is conscious of x, and any sensible interpretation identifies that y with the 'someone' in the right-hand side of (7). Thus:

(8) For a mental state x, $\exists y$ [y is conscious of x if and only if y has a thought about/perception of x] ,

which can be generalized into:

(9) $\forall x \, \exists y$ [y is conscious of x if and only if y has a thought about/ perception of x].

And that—properly understood—is all that (2) means.

The rock in (3) is either agent-conscious or patient-conscious. It cannot be the first because (1) and (2) turn out to be about patient-consciousness, so if (3) would be about agent-consciousness, then it would not follow. Therefore, (3)—like (1) and (2)—is about patient-consciousness:

(10) A rock is patient-conscious if and only if someone has a thought about/perception of that rock.

And from (4) and (10)—or from (9)—it then follows that someone

---

[3] Of course, the acceptability of (7) depends on the acceptability of (1) as a summary of HO theories, and as mentioned above, there are other problems with (1).

is conscious of a rock if and only if that person has a thought about/ perception of that rock. If 'consciousness' in (3) is patient-consciousness, then that is all that (3) means. But there is nothing objectionable to (3) understood as such (aside perhaps, from the fact that we usually do not use the adjective 'conscious' in this way in English).

The enthymematic conclusion of PoR depends on the suggestion that HO theories imply that rocks are agent-conscious (and thus that rocks have consciousness and can be conscious of things), but that suggestion is based on an equivocation fallacy. If (1) is to be a representation of the central claim of HO theories, then it is about patient-consciousness, and then no conclusion about agent-consciousness can follow. If the equivocation is avoided, then either PoR misrepresents HO theories (and thus is a straw man), or it states (the rather obvious fact) that someone is conscious of a rock if she has a thought about that rock or perceives that rock. Neither is problematic for HO theories. Therefore, there is no Problem of the Rock.[4]

Lajos Brons
Nihon University and Lakeland University
mail@lajosbrons.net

References

Byrne, Alex. 1997. Some like it HOT: consciousness and higher-order thoughts. *Philosophical Studies* 86(2): 103–29.

Gennaro, Rocco. 2005. The HOT theory of consciousness: between a rock and a hard place? *Journal of Consciousness Studies* 12(2): 3–21.

Goldman, Alvin. 1993. Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition* 2: 364–82.

Rosenthal, David. 1997. A theory of consciousness. In *The Nature of Consciousness: Philosophical Debates*. Edited by Ned Block, Owen Flanagan and Güven Güzeldere. Cambridge: Bradford. 729–53.

Stubenberg, Leopold. 1998. *Consciousness and Qualia*. Amsterdam: John Benjamins.

Van Gulick, Robert. 2000. Inward and upward: reflection, introspection, and self-awareness. *Philosophical Topics* 28: 275–305.