Commentary on Radu J. Bogdan (2012) "Self-Consciousness: Executive Design, Sociocultural Grounds." *The Fourth Consciousness Online Conference*.

Kyle Ferguson, Cognitive Science and Philosophy, CUNY Graduate Center

17 February 2012

In his interesting contribution to this conference and in his recent book, Our Own Minds (2009), Radu Bogdan offers an executive account of a psychological phenomenon that he calls *self-consciousness*. As he uses the term, 'self-consciousness' abbreviates "consciousness of self-to-target relatedness" (§2, p. 3). Bogdan advances four main and two subsidiary hypotheses. The main hypotheses are the following:

(H1) *Self-consciousness has an executive design.* Self-consciousness develops historically and ontogenetically "out of the joint work of a network of executive abilities" (§1, p. 1). Here, Bogdan is thinking of a suite of high-level executive abilities realized in the prefrontal cortex. The suite includes means-end initiative, top-down attention, monitoring and control over sensory and perceptual states, inhibition of current perception and motivation, the broadcasting of information from one mental faculty to another, multitasking, deliberate memory search, mental rehearsal, deliberate metacognition, and autobiographical recall (§4, p. 5);

(H2) *Children's intuitive-psychological understanding develops asymmetrically; first in a third-person mode, later in a first-person mode.* Our intuitive psychology, which Bogdan equates with Theory of Mind (ToM) or mindreading, first "regulates one's mental and behavioral relations to others and later oneself" (§1, 1). As it develops and is deployed, our intuitive psychology recruits the aforementioned executive abilities in response to a host of social-environmental pressures (ibid.);

(H3) *Children's self-consciousness develops asymmetrically; first outwardly, later inwardly.* Building on the asymmetric development of intuitive psychology (H2), human self-consciousness first develops outwardly or extrovertly—toward others' minds and behaviors—and later develops inwardly or introvertly—toward one's own mind and actions (§1, p. 1);

(H4) *Phenomenality, construed as conscious sensory experience, requires self-consciousness*; *but when phenomenality is construed as unconscious sentience, there is no such requirement.* Self-consciousness's executive structure, identified in (H1), is "a prerequisite and platform for the phenomenal (what-it-is-like) manifestation of consciousness" (§1, p. 1). But unconscious sentience—e.g., unconscious sensory states observed in patients with blindsight, ordinary folk's lapses into absent mindedness, and nonhuman animals' perceptual interactions with their environment—neither involves nor presupposes self-consciousness.

Bogdan's subsidiary hypotheses are as follows:

(SH1) *Self-consciousness is unique to a remarkably small number of species, those whose executive abilities are of the right level of complexity and whose environmental pressures are of the right sort*. Because it is involves complex and diverse mental machinery, self-consciousness is "likely to be a rare commodity on the evolutionary market," enjoyed only by "intensely sociopolitical species" whose members are exposed to the social-environmental pressures mentioned in (H2) (§1, p. 2);

(SH2) *When construed as mere sentience, phenomenality is widely distributed throughout the animal kingdom*. Still, sentient creatures "are likely to develop conscious branches only when engaging the relevant executive abilities" (§1, p. 2).

Before exploring any of these hypotheses in greater detail, let us consider Bogdan's methodology. He reaches the aforementioned hypotheses by adopting a "data-driven strategy," which he contrasts with methodologies relying on first-person intuitions about phenomenality or the deliverances of introspection. The data-driven strategy involves two steps. First, he identifies robust correlations between (*a*) empirical data from neurological, developmental, and clinical studies, and (*b*) the presence, absence, or degrees of consciousness. Second, he attempts to explain these correlations, offering an "evo-devo" account of why those correlations exist.

We should certainly embrace the empiricist spirit of Bogdan's approach. We would do well to refrain from relying upon the presumed accuracy of private, privileged, first-person access to our own minds. A science of consciousness seems viable only if we adopt *something like* Bogdan's methodology. But we find serious cause for pause when we reflect upon one side of the desired correlations; viz. (*b*) above. When we attribute self-consciousness to a subject—or deny such attributions—what independent reasons do we have for doing so? What licenses our saying that self-consciousness is present, absent, or diminished, in some given case? And how do we do this independent of the empirical findings with which we are correlating those ascriptions? To put the query another way, if the correlations' left-hand sides—(*a*) above—contain empirical findings, then what is supplying the right-hand sides?

Bogdan seems to think that our commonsense, folk-psychological notions of self-consciousness supply the correlations' right-hand sides. And we discover the folk's psychological commitments by observing how they tend to speak of self-consciousness. Bogdan observes that "we talk of being conscious of something or conscious that

something is the case," and that "it is hard to think of cases where [we would say that] one is simply conscious (full stop) and one's consciousness is contentless" (§2, p. 3). Because we conceive of consciousness in this manner, it follows, for Bogdan, that consciousness is *necessarily* intentional or target-related and that consciousness is *necessarily* self-consciousness or "consciousness of self-to-target relatedness" (*ibid.*).

I agree with Bogdan that our folk-psychological theorizing plays an important role in scientific investigation of the mind. But I don't think it plays that role in the way he supposes. Rather than serving as the source of necessary truths about the mind, rather than supplying the necessary, inescapable meanings of folk-psychological terms, folk psychology provides an initial, pre-reflective theory about the mind. Following David Lewis, we can't assume that all of our folk-psychological truisms will turn out to be true or survive naturalization, and we should with great humility approach the task of sorting the true—or viable—ones from the false.[1] In other words, it is possible for folk psychology to posit certain functional roles that, as it turns out and as we discover empirically, are not occupied by any neuroscientific posits.

One way of using folk-psychological concepts amounts to nothing more than conceptual analysis; another way amounts to the Lewisian-functionalist approach. I prefer the latter. But it appears that Bogdan prefers the former when he claims that certain truths about self-consciousness necessary follow from our folk-psychological concepts or from the OED's definitions of certain terms (§11, pp. 14-15). To put this

[1] See David Lewis, "Psychophysical and Theoretical Identifications," Australasian Journal of Philosophy, 50 (Dec. 1972): 249-58; "Ramseyan Humility," in Conceptual Analysis and Philosophical Naturalism, eds. David Braddon-Mitchell and Robert Nola, Cambridge, Mass.: MIT Press, 2009, Chapter 9.

objection into the form of a question, how is this aspect of Bogdan's approach distinct from conceptual analysis?

Setting the worry about methodology aside, let us turn to the correlations Bogdan identifies. We should recall that Bogdan's aim in identifying these correlations is to give empirical support for the executive design he assigns to self-consciousness. I will limit my comments to his analysis of blindsight and of the developmental literature.

First, Bogdan appeals to the phenomenon of blindsight. While blindsight patients report that they do not see, their nonverbal behavior suggests that they actually can see: they respond above chance to visual stimuli in the blind field. A person with blindsight, for Bogdan, "is not conscious of the scene in the blind field and of seeing or being visually related to the scene" (§3, p. 4). Rather, the blindsighter "unconsciously registers, processes and responds to perceptual stimuli in the blind field" (*ibid.*). For Bogdan, subjects with blindsight behave as they do by virtue of low-level executive machinery, which responds to unconscious perceptions of visual stimuli. Normal subjects can consciously perceive visual stimuli by virtue of their high-level executive machinery's access to information in the visual system. Thus, blindsight shows a strong correlation between diminished self-consciousness (a lack of it with respect to visual experience), on the one hand, and diminished involvement of high-level executive abilities. Comparing blindsighters to normal subjects, Bogdan concludes that "a high-level executive machinery correlates with self-consciousness but a low-level one need not" (§4, p. 5).

Second, Bogdan appeals to decades of research in developmental psychology to establish correlations between high-level executive abilities and self-consciousness.

While this appeal is largely an effort to substantiate (H1), it is also an attempt to underwrite (H2) and (H3). Bogdan observes that until the age of four, children's mentation is "mostly online, perceptuomotor, input-driven, and situated in the here and now of perception, motivation and action" (§5, p. 6). In their interactions with the world, these children are confined to communicational, affectuomotor and perceptuomotor modalities, where "the *motor* component and its world-directedness are essential" (*ibid.*). Such minds, Bogdan claims, rarely involve higher-level processing or the use of high-level executive abilities. And even when such executive abilities are deployed, they are externally oriented. During this time, children's memories are neither self-oriented nor autobiographical, introspection is not yet in place, and "their intuitive psychology is about other minds, not their own" (*ibid.*). During an extrovert phase of self-consciousness in early ontogeny, children's developing intuitive psychology "recruits, orchestrates and deploys in its operation a variety of high-level executive abilities that end up as a platform for self-consciousness" (§8, p. 11). Subsequently, children develop inwardly directed executive machinery, which allows them to turn their intuitive-psychological understanding of other minds toward their own minds. It is not until after the age of four or five that children develop the set of executive abilities involved in introvert self-consciousness and turn that machinery inwards. This developmental milestone, Bogdan suggests, is likely related to the major growth in children's prefrontal cortexes and their entering new sociopolitical environments during this phase of life.

Taken together, these findings are cited as support for Bogdan's executive account of self-consciousness. And the developmental data have the additional roles of

supporting Bogdan's hypotheses concerning self-consciousness's asymmetric development (H2 and H3) and partially supporting his "evo-devo" explanation the correlations identified in (H1).

I will not comment on Bogdan's proposed *explanation* of self-consciousness. Instead, I will attempt to put pressure on—or raise some questions concerning—his *description* of the phenomenon. We must evaluate a theorist's description of the phenomena they seek to explain prior to evaluating the explanations they offer. I will suggest that Bogdan may be conflating distinct phenomena when he brings them under the single heading of 'self-consciousness', and it seems that this conflation is directly related to ambiguities in the phrase "self-to-target relatedness"—that of which one is conscious when one is self-conscious in Bogdan's sense—and to ambiguities in our everyday use of the term 'consciousness'.

Consider the following scenario. Suppose we have a savvy blindsight patient. In a debriefing session following our experiment, we inform her that she scored above chance. Being savvy in her reasoning about probabilities, the patient concludes that she is seeing the visual stimulus after all. That is, she reasons about herself in much the same way that researchers reason about her. Both subject and researchers are led to infer that the subject is visuoperceptually related to the stimulus in the blind field.[2]

Is this a case of self-consciousness or not? Recall that Bogdan defines 'self-consciousness' as "consciousness of self-to-target relatedness" (§2, p. 3). The savvy

[2] It is worth noting that parallels exist between the case of the savvy, debriefed blindsight patient and a well known case described by John Perry in "The Problem of the Essential Indexical," Nous XIII, 1 (March, 1979): 3-21. Also, the inferential process I have in mind is in some ways similar to Gilbert Ryle's remarks on self-knowledge and introspection in Chapter VI of The Concept of Mind (London: Barnes & Noble, 1949).

blindsighter applies her intuitive-psychological understanding to her own behavior, forming the thought, “I am seeing the red stimulus.” A normal-sighted subject, in cases of conscious vision, forms the thought the thought, “I am seeing the red stimulus.” Plainly, the two cases are indistinguishable in terms of their respective intentional states’ contents. What, then, on Bogdan’s view, is the difference between these cases? In each cases, we have (1) an intentional state whose content includes the self-to-target relation, and (2) the use of high-level executive abilities in generating (1). In other words, what is the difference between introverted self-consciousness, on the one hand, and self-directed yet extroverted self-consciousness, on the other? As we saw above, the difference will be found in neither the intentional contents of the two states, nor the involvement of high-level executive abilities.

We might anticipate Bogdan’s response to this question. He might say that in the case of the savvy blindsighter, the extrovertly self-conscious subject is operating on information readily available to observers or from the third-person perspective. Normal-sighted, introvertly self-conscious subjects, on the other hand, operate on information or representations available from the first-person perspective. But what exactly does this information consist in? What representations—or modes of representing—are uniquely available to subjects from their first-person perspectives or uniquely involved in introverted self-consciousness?

To put the above point another way, let us move on from blindsight and towards the developmental story. We can grant that around the ages four and five, children’s prefrontal cortexes grow dramatically and their socio-political environments shift in important ways. But what exactly, in terms available at the functional level of describing

psychological systems, goes on during this period to move children from the extrovert phase to the introvert phase of self-consciousness? To what are they turning their executive machinery when they turn it inwards? What representations or modes of representing are newly available to children at this point in time?

These questions must be answered if we are to appreciate and understand the phenomena Bogdan seeks to describe and explain. In order to answer these questions, we must identify the exact notion of 'self-to-target relatedness' Bogdan has in mind when he defines 'self-consciousness' as "consciousness of self-to-target relatedness." There are at least five ways of construing the relation, of construing *that of which one is conscious when one is self-conscious*. Let '*R*' be the relevant relation, '*o*' the relevant organism (e.g., a six-year-old named Oscar), '*s*' the relevant stimulus (e.g., a strawberry), and '*ms*' the relevant mental state (e.g., the visual perception of the strawberry):

(1) $R$ <*o*, *s*>

(2) $R$ <*o*, *ms*>

(3) $R$ <*ms*, *s*>

(4) $R$ <*o*, *ms*, *s*>

(5) $R$ <*o*, $R$ <*ms, s*>>

There are various ways of representing the self-to-target relation. Each way exploits ambiguities in 'self' and 'target', or is owed to the fact that there are the various descriptions we can use to refer to the self, its mental states, and environmental stimuli. Does introvert self-consciousness involve some particular mode of self-reference, or is any sort of self-reference sufficient? If some particular mode, what is its nature? A

similar question can be asked about the references to one's targets. Must these references involve *explicit* representation of mental states—or stimuli—as such, or will *implicit* reference suffice? Finally, when subjects are introvertly self-conscious, how do they represent or characterize the relation (*R*, above) between themselves and their targets?

In the end, it is unclear whether 'self-consciousness' refers to a category of mental states or to subset of states therein. Sometimes 'self-consciousness' seems to include any mental state involving self-reference and asserting a relation between that referent and a mental state; other times it seems to pick out a narrower set of states in that category. The waters here are muddied by Bogdan's later remarks, where self-consciousness turns out to not require occurent mental states at all. Instead, it amounts to a "dispositional readiness, always on, when awake, to allocate the right executive resources to handle the task-directed mental states in the right patterns with the right intensity" (§12, p. 15).

I have been suggesting that there is a lack of clarity with respect to the phenomenon that Bogdan seeks to explain. I'll conclude by suggesting that this likely results from Bogdan's running together distinct senses of 'consciousness', and I'll point to a way achieving clarity where it is needed.[3] Following David Rosenthal, we can distinguish between three senses or everyday uses of 'consciousness': *creature consciousness*, *transitive consciousness*, and *state consciousness*.[4] Sometimes we say that we have regained consciousness, and all that we mean is that we are now

[3] This conflating is especially apparent in §12.
[4] See, e.g., David M. Rosenthal, Consciousness and Mind (New York: Oxford University Press, 2005, Chs. 2, 4, and 7).

awake, responsive to environmental stimuli, or that the anesthesia has worn off (creature consciousness). Other times we say that we are conscious of something or conscious that something is the case (transitive consciousness). It should be noted that this transitive sense is the sense of 'consciousness' Bogdan notes in §2 (p. 3). When we attribute transitive consciousness, we attribute nothing more than some intentional state whose content is simply the object or state of affairs of which one is conscious. Finally, there is the sense of 'consciousness' that refers to *state consciousness*. The state-consciousness problem is the problem of identifying and explaining what it is for a mental state to be conscious as opposed to unconscious. And the higher-order-thought theory Rosenthal develops is a proposed solution to the state-consciousness problem.[5] I think we can better understand Bogdan's theory by getting clear on the problem that it is supposed to solve. What exactly is the self-consciousness problem? And how does the self-consciousness problem compare to state-consciousness problem?

[5] As it turns out on a higher-order-thought theory of state consciousness, when I token a HOT that I am seeing a red stimulus, I am transitively conscious of my perceptual state. But transitive consciousness of oneself or of one's mental state is not sufficient for state consciousness on the HOT view. On that view, there is a second necessary condition according to which the HOT seems subjectively immediate and non-inferential.