



Reply to Legrand: Content from the Inside Out

Thomas Metzinger
Philosophisches Seminar
Johannes Gutenberg-Universität Mainz
D-55099 Mainz
www.philosophie.uni-mainz.de/metzinger
metzinger@uni-mainz.de
© Thomas Metzinger

PSYCHE 12 (4), August 2006

Reply to: Legrand, D. 2005. Transparently Oneself, *Psyche* 11(5).

Keywords: Self, phenomenal self-model, phenomenal content, body, embodiment, phenomenal transparency, phenomenal opacity, dynamical sensori-motor coherence, inwardness

In the current debate, very few people have penetrated as deeply into the self-model theory of subjectivity and have developed such a scholarly expertise on the project as a whole (including all its difficulty) as Dorothée Legrand has done. In the last sentence of her commentary, Legrand alludes to the ugly consequences I have to face after calling the book *Being No One*: I am suddenly confronted with people from all over the world who are stomping their feet on the ground like stubborn children, claiming that *they* definitely are *someone*, and that *they* definitely have a self. In 1993, I published a German precursor to BNO, titled *Subject and Self-Model—The Perspectivalness of Phenomenal Consciousness Against the Background of a Naturalist Theory of Mental Representation*. From this title alone, it can be seen that I was not interested in demolishing classical theories of the self as a metaphysical substance, but rather in an empirically informed representationalist approach to the phenomenal self. In particular, I wanted to get a grip on the relationship between the epistemic subject and the phenomenal self it uses in the attempt to gain self-knowledge. The mistake I made, and to which Legrand alludes at the end of her brilliant commentary, was assuming that for a predominately American readership, such a title would be much too “wordy.” Now I am paying the price – the price for having diverted many of my readers’ attention to discussions about how our folk-phenomenological intuitions about selfhood can somehow be rescued. The title *Being No One* has distracted a lot of attention that would probably be better focused on the subtle and difficult issues associated with the attempt to construct a tenable theory

about the interaction between self-knowledge and self-experience. But Dorothee Legrand *cannot* be distracted.

On page 4, Legrand writes that, in my terminology, mental content is phenomenal “when it does not depend on the existence of the represented object.” It is true, with the concept of “phenomenal content” I aimed at exactly that aspect which is the same for a veridical perception of an apple and an (undetected) hallucination of an apple. Intentional content has epistemic status. Phenomenal content hasn’t. Phenomenal content is that aspect of intentional content that is neither true nor false, because it solely determines how things *appear* to you. Phenomenal consciousness is *appearance*. And the phenomenal *self-model* determines how you *appear to yourself*. Intentional content, on the other hand, can be part of a representation or part of a misrepresentation. It can be ascribed from the outside, from a third-person perspective. Phenomenal content, so far, is exclusively accessed from the “first-person perspective” (and one of the central goals of SMT is to develop a theory of what such a perspective is in the first place). The intentional content active in a given system is frequently determined by external factors, meaning that its reference may be linked to social facts and that its epistemic status may depend on the existence of external target objects, for instance the apple in front of your eyes. Phenomenal content, however, supervenes locally: it is fully determined by functional properties internal to the system, in our case by functional properties of its minimally sufficient neural correlate. The most serious misunderstanding, however, would be the assumption that the distinction between “intentional” and “phenomenal” content is *exclusive*: according to SMT, phenomenal content is a *special kind* of intentional content, one satisfying the additional constraints described in chapter 3. One of the many ways in which it is a *special kind* of intentional content is that, ontologically, it is determined by spatially and temporally internal properties of the central nervous system. The reason for this is simple: The functional mechanisms responsible for realizing the additional properties (e.g., global availability of information for attention, cognition, and rational motor control or the integration into a virtual window of presence) are all physically internally realized by the human brain, and only by the human brain. So we get: (i) not all intentional content supervenes locally; (ii) all phenomenal content does supervene locally; (iii) phenomenal content is a special kind or subset of currently active intentional content, generated by additional, internally realized mechanisms. So it is not entirely true that intentional content depends on the existence of the represented object—in its *epistemic status*, it depends on its existence. And it is not entirely true (see p. 4) that phenomenal content does not depend on the existence of the represented object. True, phenomenal content *per se* has no epistemic status. In forming this concept, we were not interested in knowledge at all, but in the deep structure of appearance. Nevertheless, please note how, under a *genetic* perspective, it would be hard to understand the evolution of consciousness without the evolution of representational states: if there had never been apples in the environment of our biological ancestors, the representational resources to form the respective form of intentional content would never have evolved. And if there had never been any unconscious intentional states of this type, we would not possess phenomenal apple-experiences today. From a teleofunctionalist perspective, under the adaptivity constraint (BNO: section 3.11), the phenomenal experience of apples actually did depend on their past experience. So I am not saying that phenomenal content is an *entirely* different kind of mental content. What I am saying is that it is a particular

aspect of only a proper subset of the intentional content active in an organism at a given time.

On page 6, referring to similarities between offline and online hallucinations, Legrand writes that I disregard a crucial difference, namely that “the representational carrier represents itself in the case of a simulation, while it represents something else than itself (the world) in the case of a representation.” This may be a misunderstanding, resting on an equivocation between “representation” and “simulation” used as epistemological concepts, as was the case in chapter 2, and the *phenomenology* of simulation and of representation. I do not really understand how a representational carrier could represent itself *as a carrier*, but there may be some formal solution to this question abstracting from the temporal dynamics and the whole-system context. In any case, phenomenal opacity, the conscious experience of the representational nature of an ongoing phenomenal state, has to do with the construction process, with earlier processing stages becoming globally available. But this is an introspective process, a process of higher-order representation, and not a process by which a carrier represents itself. This, in turn, is a functional-level hypothesis about the phenomenal-level property of opacity. I am not sure if this is the functional difference Legrand actually had in mind, but it doesn’t touch the general epistemological claim that, in our own case, and as regards the temporal features of intentional content, every representation is also a simulation, simply because it is a time-consuming process that never brings us into direct and immediate contact with our present environment. The internally constructed Now is not the physicist’s knife-edge, but rather William James’ saddleback.

But there is a deeper problem, for which I have no ready-made solution, and which is also touched upon in my reply to Gallagher’s commentary. Could phenomenal content appear in the absence of intentional content? Not if phenomenal content is an aspect of a proper subset of the intentional content currently active in a given system – the pre-existence of intentional content would then be a necessary condition for the additional aspect of phenomenality to come into being. But let us ask: Is “pure appearance” a conceptual possibility? SMT, of course, is concerned with ecologically valid situations, and with systems possessing a true telefunctionalist analysis. But what about hypothetical situations, like Shaun Gallagher’s “brain-in-a-vat embodiment”? In the spirit of Daniel Dennett’s intentional-systems theory, we might want to say that a brain in a vat has so many false beliefs *de se* that it certainly cannot count as rational anymore, and we might even want to go so far as to say that, given its current state, it has *no* representational properties whatsoever. Even though its internal states may have had a biological function for its ancestors or for a fully embodied, situated organism, the brain in a vat is simply a system possessing certain physical and functional properties—but no intentional content. On the other hand, it is overwhelmingly plausible to say that our brain in the vat would experience *something*. As noted above, all the empirical evidence about direct brain stimulation, dreams and hallucinations clearly points into this direction. According to the original thought experiment, our brain in a vat would still realize those functional properties on which the phenomenal content supervenes locally. But why would this still be a form of “content,” if there were no representational properties to be constrained? Can we allow for systems possessing physical, functional, and phenomenal properties, but no *representational* properties? I think that the empirical *prima facie* plausibility of 3E (see my reply to Gallagher’s commentary) and other phenomenal

properties in weakly representational systems constitutes a good argument for their reducibility in neurofunctional terms. I also admit that I have no official position on the possibility of non-representational phenomenal character or on the possibility of phenomenal properties in the strict absence of representational properties. The question would be whether such situations actually do occur in ecologically valid, nomologically possible scenarios, or only in logically possible worlds like those described by our thought experiment. Deciding the issue would ultimately depend on what our epistemic interests and relevance criteria actually are.

On page 6, Legrand demands that I at least indirectly acknowledge a form of externalism for phenomenal experience. Now we can see how this could be done: in real-world systems, the existence of phenomenal content is parasitic on the occurrence of intentional content, which must be individuated in an externalist fashion. If this content satisfies the conditions described by SMT, it can be elevated to the level of conscious experience. Phenomenal content still supervenes locally, but due to its transparency, it endows us with a robust, false intuition, namely that it is *directly* determined by external factors like the existence of perceptual objects. This false intuition then carries over into our attempts at developing a theoretical understanding, generating all the well-known conceptual confusions and difficulties.

In section 2, Legrand frequently equivocates between “phenomenal” and “phenomenological.” “Phenomenal” refers to concrete states, “phenomenological” refers to statements or theories about such concrete states. Philosophically this is, of course, a major difference, but as the standard usage of these terms is incoherent and fluid in the community itself, this certainly is not a major point of criticism. It may, however, sometimes play a role. I would also not say that phenomenal experience is *experienced* as transparent (p. 9). First, not all phenomenal experience is transparent, it is just that, for fully transparent phenomenal states, content properties are the only introspectable properties. Compare the phenomenology of conceptual thought, for instance, the conscious experience going along with an episode of philosophical reasoning. Here, we are aware of the fact that we are currently operating with mental *representations*, which could be true or false. We are subjectively aware of the ongoing construction process and the different stages of the *formation* of thoughts, of the different stages of disambiguating them or associating them with other thoughts, embedding them into a context, etc. That is, earlier processing stages—non-intentional properties—are actually available for introspective attention all the time. This is why conscious thought is phenomenologically opaque, we experience it as made, and not as given. But for fully transparent phenomenal states (like sensations), content properties are the only introspectable properties. This fact in itself does not “reveal” (p. 9) anything about the true nature of experiential states. Though I mostly agree with Legrand’s careful description of the terrain, I must admit that I am not quite sure how to frame the possibility of “metaphenomenal” content in a conceptually precise way: Can we really experience experience as such, is there metaphenomenal content in the same sense that metarepresentational content exists? A first-order phenomenal content (e.g. the sensation of blue) would then become the intentional object of a higher-order mental state, which in addition would satisfy the constraints for phenomenality. Would this process add anything to the first-order phenomenal content on the level of second-order phenomenal content? For instance, would the first-order state now lose its transparency, because it would suddenly be

experienced *as representational*? In section 3, Legrand presents us with a very interesting and stimulating idea: “Inwardness implies that experience is experienced” (p. 11). In BNO, I have taken care to distinguish a number of different notions of “inwardness.” Legrand now introduces her own conception and interestingly ties it to the notion of phenomenal transparency.

First, note that it is not strictly true that the “transparency/opacity continuum is an aspect of reflective consciousness” (p. 12). I have presented examples of visual pseudohallucinations that can take place in what is experienced as the external, objective world (e.g., as abstract geometrical patterns on a wall) and involve no form of reflexivity. A conceptual distinction used in neurology and psychiatry, which is of great importance to philosophers, is the one between “pseudohallucinations” and “complex hallucinations”. The patient experiencing pseudohallucinations *knows* that he is currently hallucinating, whereas the patient suffering from complex hallucinations is severely deluded, he does not understand or grasp the fact that he is actually hallucinating. If I understand correctly Legrand’s claim about transparency not meaning invisibility, this would mean that a visual pseudohallucination would not only be an example of the emergence of phenomenal opacity in the system’s model of the external world, but it would at the same time be an example of the emergence of *inwardness*: At the very moment I lose naïve realism, at the moment I discover that something must be wrong with my visual perception, I would also discover that something must be wrong with *myself*—I would appropriate a part of the process of sensory representation. This may in fact be a highly interesting and very subtle phenomenological observation concerning opacity. But it would also lead to the conclusion that in our standard example for *globalized* phenomenal opacity—the lucid dream—the dreamer’s experience should be one of total inwardness, an experience of “all of this is really happening inside of myself.” And this is not the case. The lucid dreamer may think a *thought* with this content, but the phenomenological profile of this state class is clearly different. This means that in the future, a lot of careful conceptual work has to be done in developing and distinguishing different notions of “phenomenal transparency,” “phenomenal opacity,” and “inwardness.”

Let us now turn to Legrand’s own positive conception. At the beginning of section 4, she makes a mistake many people make (see for instance Zahavi 2006, p. 1, 103; Gallagher, this issue), namely misconstruing my claim as saying that the self is an “illusion.” I am concerned with the prereflexive, preconceptual, subsymbolic, and subpersonal roots of the phenomenal self. On this level, neither truth and falsity, nor knowledge and illusion exist in any philosophically interesting sense. I must admit that I have not understood what exactly the argument is for saying that the claim “SMT-systems exist” *means* “selves exist.” Again, this looks like a simple *petitio* to me, presupposing what remains to be shown by an independent argument. If one reads on, however, one immediately realizes that Legrand is simply presenting an *alternative* notion of what a self is, and one that has many interesting aspects. It is the prereflexive “structure of phenomenal experience” itself (p. 15) and in a clearly Kantian vein, it is something that “is not invisible, but transparent in that it reveals the world by hiding itself” (p. 15). This is a very metaphorical and beautiful way of speaking, but it also helps to make a classical point: self-consciousness may be a necessary condition for the possibility of epistemic experience. As opposed to the classical intuition in transcendental philosophy, these structural features would not be strictly non-phenomenal. What

Legrand envisions are those fundamental aspects of self-experience which do not form an explicit form of mental content, but which are still implicitly present as the deep structure of phenomenal experience.

I disagree that the descriptions of the phenomenal content associated with an active PMIR on page 16 of the précis of BNO necessarily imply taking oneself as an object of attention. As clearly shown in the figures of the précis, the idea is that system-object relationships can be experienced and available for attention without actually being accessed introspectively. The quotation of John Perry's argument is inappropriate in this context, because the whole point is not about how to express knowledge, but about a specific form of phenomenal content, namely the PMIR. This phenomenal content needs to be expressed in the way I have done.

Nevertheless, Legrand's own strategy of avoiding the classical phenomenological fallacy by "dynamicizing" the phenomenal self is interesting and well worth pursuing further. In a recent paper, Legrand has offered a view of the body as "dynamical sensorimotor coherence" (Legrand 2005: 115) and anchored her own conception of pre-reflective self-awareness in this idea. The most interesting claim here, and one to pursue in the future, is that pre-reflective, bodily self-consciousness is not private, but "open." If fleshed out, would this idea connect to Legrand's own notion of inwardness in an interesting way?

References

- Legrand, D. (2005). The bodily self. The sensorimotor roots of pre-reflective self-consciousness. *Phenomenology and the Cognitive Sciences* 5: 89–118.
- Metzinger, T. (1993). *Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation*. Paderborn: Schöningh.
- Zahavi, D. (2006). *Subjectivity and Selfhood*. Cambridge, MA: MIT Press.