The Journal of Mind and Behavior

Vol. 35 No. 3 Summer 2014

ISSN 0271–0137

The Journal of Mind and Behavior (JMB) is dedicated to the interdisciplinary approach within psychology and related fields. Mind and behavior position, interact, and causally relate to each other in multidirectional ways; JMB urges the exploration of these interrelationships. The editors are particularly interested in scholarly work in the following areas: □ the psychology, philosophy, and sociology of experimentation and the scientific method □ the relationships among methodology, operationism, and theory construction □ the mind–body problem in the social sciences, psychiatry and the medical sciences, and the physical sciences □ philosophical impact of a mind–body epistemology upon psychology and its theories of consciousness □ critical examinations of the DSM–biopsychiatry–somatotherapy framework of thought and practice □ issues pertaining to the ethical study of cognition, self-awareness, and higher functions of consciousness in nonhuman animals □ phenomenological, teleological, existential, and introspective reports relevant to psychology, psychosocial methodology, and social philosophy □ historical perspectives on the course and nature of psychological science. We typically do not publish empirical research. The Journal also recognizes the work of independent scholars.

JMB is based upon the premise that all meaningful statements about human behavior rest ultimately upon observation — with no one scientific method possessing, a priori, greater credence than another. Emphasis upon experimental control should not preclude the experiment as a measure of behavior outside the scientific laboratory. The editors recognize the need to propagate ideas and speculations *as well as* the need to form empirical situations for testing them. However, we believe in a working reciprocity between theory and method (not a confounding), and in a unity among the sciences. Manuscripts should accentuate this interdisciplinary approach — either explicitly in their content, or implicitly within their point of view.

JMB offers a publication outlet on a quarterly basis. The Journal publishes one volume per year in the following sequence of issues: Winter, Spring, Summer, and Autumn. There are no submission fees or page costs for accepted manuscripts. JMB is a peer-reviewed, refereed journal, and all decisions will be made by the assessing editors, associate editors, and chief editors. Commentaries and responses to individual articles and reviews are welcome. Authors subscribing at the time of manuscript submission are eligible for reduced subscription rates (see below).

All manuscripts should follow the style and preparation presented in the *Publication Manual of the American Psychological Association* (sixth edition, 2010). Particular attention should be paid to the citing of references, both in the text and on the reference page. [Note exceptions to APA style: JMB uses *no* ampersands or city/state abbreviations in referencing; the Journal uses three levels of headings: level 1, level 3, and level 4, see pp. 113, 114, 115 from the fifth (2001) edition APA Manual.] Authors requesting blind review must specify and prepare their manuscripts accordingly. Manuscripts may be sent to the Editor either by e-mail to jmb@maine.edu or by post (one copy) to:

Raymond Chester Russ, Ph.D., Editor
The Journal of Mind and Behavior
Department of Psychology
The University of Maine
5742 Little Hall
Orono, Maine 04469–5742
Tel. (207) 581-2057

Yearly subscription rates are $32.00 for students or hardship; $35.00 for past/present JMB authors or for those submitting manuscripts; $46.00 for individuals; $185.00 for institutions. Air mail rates upon request. All back issues are available and abstracts are located at www.umaine.edu/jmb. For detailed information contact our Circulation Department at The Institute of Mind and Behavior, P.O. Box 522, Village Station, New York City, New York 10014; Tel: (212) 595-4853.

continued inside back cover

The Journal of Mind and Behavior

Editorial Board

Assessing Editors

The Journal of Mind and Behavior

Vol. 35 No. 3

Summer 2014

Library of Congress Cataloging in Publication Data

The Journal of mind and behavior. – Vol. 1, no. 1 (spring 1980)–
– [New York, N.Y.: Journal of Mind and Behavior, Inc.]
c1980–

1. Psychology–Periodicals. 2. Socialpsychology–Periodicals. 3. Philosophy–Periodicals. I. Institute of Mind and Behavior
BF1.J6575 150'.5 82-642121
ISSN 0271-0137 AACR 2 MARC-S

The Journal of Mind and Behavior
Summer 2014 Vol. 35 No. 3

CONTENTS

The Journal of Mind and Behavior
Summer 2014, Volume 35, Number 3
Pages 107–132
ISSN 0271–0137

Knowing How it Feels: On the Relevance of Epistemic Access for the Explanation of Phenomenal Consciousness

Itay Shani

Kyung Hee University

Consciousness ties together knowledge and feeling, or sapience and sentience. The connection between these two constitutive aspects — the informational and the phenomenal — is deep, but how are we to make sense of it? One influential approach maintains that sentience ultimately reduces to sapience, namely, that phenomenal consciousness is a function of representational relations between mental states which, barring these relations, would not, and could not, be conscious. In this paper I take issue with this line of thought, arguing that neither of these salient aspects of consciousness reduces to the other. Instead, I offer an explanatory framework which takes both sentience and sapience as ontological fundamentals and explore how they co-evolve. In particular, I argue that while epistemic access cannot generate experience from scratch it does play a crucial role in constituting an important form of higher-order experience, namely, the capacity to experience a sense of ownership over one's experiential domain.

Keywords: sapience, sentience, panpsychism

> "Herein lies the great mistake of the Cartesians, that they took no account of perceptions which are not apperceived."
>
> Leibniz
> *Monadology*

Consciousness is a multifaceted phenomenon, and the concept of consciousness is a mongrel connoting a variety of different senses (see Block, 1995; van Gulick,

This work was supported by the *National Research Foundation of Korea Grant funded by the Korean Government* (NRF 2010–220–A0001). A much earlier and considerably different version of the paper was presented on February 9, 2013 at the *Consciousness and Intentionality: Franz Brentano's Heritage in the Philosophy of Mind* conference in Salzburg, Austria. I thank the participants and organizers of this conference and in particular Johannes Brandl, Uriah Kriegel, and Elisabetta Sacchi, for their thoughtful comments. I also thank Liam Dempsey, an unknown reviewer, and the editor of this journal, Raymond Russ, for their wise advice. Correspondence concerning this article should be addressed to Itay Shani, Ph.D., Department of Philosophy, Kyung Hee University, 1 Hoegi dong, Dongdaemun gu, Seoul, 130–701, Korea. Email: ishani479@hotmail.com

2014). It is widely assumed that the puzzle of consciousness is the puzzle of fitting this multifaceted and remarkable phenomenon into the wider nexus of reality, finding it a place in nature, as it were (see, for example, Rosenberg, 2004). Yet, it is not hard to see that part of the puzzle is also internal, namely, that the challenge consists, in part, in the difficulty of bringing the various facets of consciousness into mutual accord. When the task emphasized is that of integrating consciousness with the rest of nature we ask questions such as: Can the raw feels of conscious experience be nothing but physical processes? Or, can the subjective dimension of consciousness be a part of an objective physical order? In contrast, when emphasis is laid on the internal task of bringing the various facets of consciousness into mutual accord the relevant query is of a different type, to wit: What has one aspect of conscious experience, X, to do with another aspect Y? It is this latter sort of problem which occupies me here. More specifically, my goal is to investigate the nature of the connection between two of the most fundamental features of consciousness: the phenomenal dimension of felt experience, and the cognitive dimension of knowledge and information processing.

Feeling and knowing are, without a doubt, among the most recognizable, general, and fundamental features of consciousness. When conscious, we *experience*, and there is a felt quality, a phenomenal character or "something it is like," to our experience. But, when conscious, we are also *aware* of something, which is to say that the act of experiencing is also an act of knowing, laden with epistemic qualities. Philosophers refer to these two aspects, or features, as "sentience" and "awareness" (de Quincey, 2002), "experience" and "information" (Flanagan, 1992), or, more technically, "phenomenal consciousness" and "access consciousness" (Block, 1995). It would be nice to be able to explain the exact nature of the connection between these two basic features of consciousness, assuming, of course, that the correlation between the feeling component and the knowledge component is more than mere accident. In particular, it would be nice to know whether one of these two constitutive aspects of consciousness is ontologically prior to the other, the latter being, in some sense, derivative of the former; or, if the two are mutually dependent and none is more basic than the other, to know how they contribute to each other's structure and character. In short, it is desirable to be able to answer the question what have the knowledge aspect and the feeling aspect of consciousness to do with each other.

Alas, the problem is convoluted, admitting no simple answers. For the sake of making the present discussion manageable I shall narrow down the domain of inquiry by focusing on a question which can be framed in unidirectional terms, namely: What, if any, is the explanatory relevance of the knowledge aspect of consciousness (KAC) to the feeling aspect of consciousness (FAC)?

In other words, the question is how we should understand the role of sapience in the making, or shaping, of sentience.[1]

This question can, in turn, be broken down into yet more specific components. In particular, it is useful to structure the discussion around two complementary issues. The first issue is whether or not it is possible to *derive* FAC from KAC. Clearly, if sentience is reducible to sapience then there is a very obvious sense in which KAC is explanatorily relevant with respect to FAC, namely, that the feeling aspect of consciousness is but a function of, or a specialization within, the knowledge aspect of consciousness. But if no such reduction is in the cards it becomes less clear what, if any, is the role of sapience in the making, and shaping, of sentience. Hence, in the eventuality that phenomenal consciousness cannot be derived from the purely informational components of consciousness, the next issue which confronts us is what alternative role might sapience still play in the explanation of sentience.

My goal is to articulate such a non-reductive alternative, namely, to explain in what sense sapience is still indispensable for a proper understanding of sentience even if we relinquish the hope (or the nightmare?) of reducing the latter to the former. Throughout the discussion, I shall assume that there is an important sense in which phenomenology is in the head, or, at any rate, the head and body. Those who are sympathetic to *phenomenal externalism* (see, e.g., Dretske 1995; Lycan 1996; Tye 1995), i.e., to the idea that the qualitative features of experience consist entirely in properties of the objects of experience (or, in other words, that there is nothing internal about the phenomenal character of conscious mental states) may find little to trouble them in what follows. To attempt to refute this influential position is something which I cannot do on the present occasion without sinning against bulk and thematic balance. There is enough sense, I believe, in exploring the relationships between sentience and sapience based on the traditional and still popular idea that consciousness presents us with an unflinchingly inner dimension of reality, leaving the question whether the belief in such an inner dimension is justified for another occasion.

FAC from KAC: Can Sentience be Reduced to Informational Access?

Those who are realists about phenomenal consciousness, and who take it to be a natural phenomenon, can agree at least on one thing: that consciousness

[1]Here and elsewhere in this paper I use the term "sapience" as the informational correlate to the feeling aspect of consciousness (viz., to sentience, or phenomenal consciousness). Although this term is somewhat archaic its meaning reflects accurately the knowledge aspect of consciousness. As such, it has an advantage over more frequently used terms such as "awareness," which are rife with phenomenal connotations. Thanks are due to Liam Dempsey for suggesting this term to me and for pointing to its earlier use by Feigl (1958).

as we know it, i.e., the kind of non-transitive consciousness with which we are all intimately familiar on a regular basis, is an ontological novelty. By describing it as an ontological novelty I mean simply this: that, as a natural kind, such consciousness did not exist from the very beginning of things. It came into being sometime during the course of cosmic evolution, indeed, by all accounts, relatively recently. If we were to go back in time two billion years, nothing here on Earth would be endowed with anything resembling our own states of consciousness, or those of other recently evolved intelligent species.

This much can be agreed upon not only by orthodox physicalists but also by panpsychists, neutral monists, and even absolute idealists. Thus, for example, panpsychists need not deny that consciousness as we know it is markedly distinct, qualitatively speaking, from the micro-phenomenal states which, they hypothesize, are enjoyed by unicellular organisms, molecules, etc. On the contrary, most panpsychists would agree that the differences between the micro-phenomenal and the macro-phenomenal levels are, in all likelihood, staggering.

However, attempts to describe the *nature* of this coming into being of macro-phenomenal consciousness quickly lead to wide disagreements. Physicalists typically accept a metaphysical picture according to which the antecedent physical conditions which gave rise to the evolution of macro-phenomenal consciousness, as well as the physical "building blocks" whose combination gives rise to tokens of experience, are ultimately devoid of subjectivity and sentience. In its default, aboriginal state, nature is utterly numb, lacking an interior and certainly lacking anything which remotely resembles phenomenal consciousness, or which could be considered a precursor of experience. To varying degrees, supporters of panpsychism, neutral monism, absolute idealism, and even certain versions of unorthodox physicalism all deny this basic assumption.

By affirming that nature's default state is categorically objective and insentient, orthodox physicalists commit themselves inadvertently (to the extent that they are realists about consciousness, that is . . .) to the idea that phenomenal consciousness is a completely new ontological kind, categorically distinct from, and utterly discontinuous with, anything else in nature. Such a position is often referred to as *radical emergence* (see Seager and Allen–Hermanson, 2013; van Gulick, 2001), a view which Galen Strawson describes as holding that there is nothing "about the nature of the emerged-from (and nothing else) in virtue of which the emerger emerges as it does and is what it is" (2006, p. 15).[2]

[2]This commitment to radical emergence is inadvertent insofar as many (perhaps most) orthodox physicalists are firm believers in a physically reductionist explanation of consciousness. As reductionists, they would be very reluctant to align themselves with an idea whose flavor is reminiscent of the highly non-reductive doctrines of good old British emergentism. Yet, the blatant discontinuity between a "dead," categorically insentient universe and the reality of subjective experience seems to leave the qualia realist with little choice but to affirm radical emergence.

It is possible for an orthodox physicalist to maintain that the emergence of sentience from the dim background of an utterly insentient world is a brute physical fact, the result of a dynamical configuration of processes (typically, neurophysiological ones) whose activation just happens to "switch the lights on" as it were. Indeed, I argue in the next section that such a scenario seems inevitable on the assumption that complete insentience is nature's default state. However, as a matter of fact, the neural configurations which are being identified as the underpinning correlates of conscious experience are typically portrayed as psychologically meaningful. For example, in the works of such authors as Crick and Koch (1990), Dehaene and Naccache (2001), and Edelman (1992), the identified correlates of conscious experiences are all processes which are presumed to be responsible for large-scale integration and retrieval of information, often through attentional amplification. Likewise, in Damasio's (1999) theory of core consciousness, experience is associated with meta-representations of the interactions between subject and world. In other words, the neural correlates of conscious experience are processes laden with cognitive significance corresponding to what was identified earlier as the knowledge aspect of consciousness.

That this is the case is not surprising given the robust correlation between FAC and KAC, or between sentience and sapience, but it leaves open the question just why the cognitive processes thereby identified are endowed with a phenomenal feel. In other words, there remains the question why must sapience be accompanied by sentience. Thus, while the explanatory burden may shift from the neurophysiological domain to the cognitive domain the gap in the explanation of phenomenal consciousness remains (see Chalmers, 1996). One way of approaching a solution to this problem is via what I call the *FAC-from-KAC* hypothesis, namely, the hypothesis that phenomenal consciousness is reducible to, or is a function of, representational relations between mental states which in themselves are not, or need not be, phenomenally conscious.

To recapitulate, the idea is that sentience is explainable in terms of sapience. In principle, it is possible to adopt a more modest stance: treating the correlation between sentience and sapience on a purely descriptive level. For instance, one could observe that certain forms of cognitive awareness (perhaps re-entrant signalling, or perhaps higher-order monitoring) are invariably accompanied by phenomenal consciousness and yet one can refrain from making the stronger claim that the latter results from, or is reducible to, the former. To argue in favour of the stronger thesis is to take an explanatory rather than a purely descriptive approach. This stronger thesis is, of course, more ambitious and therefore more exciting. Correspondingly, there is no shortage of bold theorists willing to pursue this explanatory project by articulating one variant or another of the FAC-from-KAC hypothesis. Below, I consider three different varieties of this ambitious agenda representing higher-order monitoring, self-representational, and thick specious present theories, respectively.

David Rosenthal, a leading advocate of the higher-order thought (HOT) account of consciousness, is a clear defender of the FAC-from-KAC hypothesis (see also Carruthers, 1996; Gennaro, 1996). On Rosenthal's (2002) view, a mental state M is conscious (i.e., non-transitively conscious) if and only if there is another mental state M* such that M* is an occurrent higher-order thought representing M as its object. Thus, it is in virtue of being represented by a HOT M* (which in itself may be either conscious or unconscious, as the case may be) that M becomes conscious. Rosenthal is aware that it is, prima facie, far from obvious that the mere fact that M is represented by M* should account for there being something it is like to be in state M. Nevertheless, he proceeds to defend just that, arguing that "being able to form intentional states about certain sensory qualities must somehow result in being able to experience those qualities consciously" (p. 413).

Rosenthal's justification of this bold claim is somewhat complicated but, in essence, it consists of the idea that our ability to be conscious of sensory qualities is contingent on our capacity for making appropriate conceptual discriminations, discriminations which are captured in the form of higher-order thoughts. The more able we are of making such conceptual discriminations, the greater the variety of sensory qualities we can experience. For example, "learning new concepts for our experiences of the gustatory and olfactory properties of wines typically leads to our being conscious of more fine-grained differences among the qualities of our sensory states" (ibid.). Conversely, he argues that the lesser the amount of classificatory HOTs at our disposal the duller and the more generic our experience becomes, such that peeling away all HOTs "would result, finally, in its no longer being like anything at all to have that sensation" (ibid.) The moral, then, is that HOTs are both necessary and sufficient for phenomenal consciousness.

Self-representational accounts of consciousness are often advanced in contrast to higher-order monitoring theories (see Kriegel, 2007), yet at least one defender of the self-representational view — Greg Janzen — shares Rosenthal's explicit endorsement of FAC-from-KAC. Janzen argues that "the phenomenal character of at least perceptual consciousness can be fully explained in terms of self-awareness, i.e., in terms of a low-level or 'implicit' self-awareness that is built into every conscious perceptual state" (2006, p. 44). On the self-representational view of consciousness, whose roots are traceable to Brentano (1874/1995), every conscious state is *bidirectional*: it is intentionally directed at (i.e., transitively conscious of) something other than itself, but in addition it also curls upon itself. Although largely *implicit* (owing to the fact that, normally, one's attention is focused on the intentional object rather than on the internal representational medium), such reflexive self-awareness is a constant presence, lurking in the background.

As mentioned above, Janzen argues that implicit self-awareness is literally constitutive of the phenomenal character of perceptual states and, possibly, of

many other conscious mental states. His account is deflationary, consisting, in essence, of the idea that the particular "what it's like" of a perceptual state P consists of the particularities of the process of perceiving x, the perceptual object, while the self-awareness component guarantees that this "what it's like?" of P is present to the cognitive subject. The result, allegedly, is that there is *something it is like for me*, the subject and the owner of P, to perceive x, and this, he argues, offers a coherent solution to the problem of phenomenal character.[3]

Finally, the FAC-from-KAC hypothesis is also espoused by Nicholas Humphrey (1992, 2000), as well as by Ralph Ellis and Natika Newton (Ellis and Newton, 2005) who, like Humphrey, pursue the subject from an action-oriented perspective in which the notion of a temporally thick present plays a prominent role. In both accounts the central idea seems to be that phenomenal consciousness, the raw feels of experience, results from the superposition of distinct temporal moments onto a unified thick specious present. This partial overlap between memorized past, occurrent present, and anticipated future allows for information to coalesce into coherent units of extended "thick" moments in which representations of past, present, and future are vividly accessed, and are recruited at the service of meaningful action guidance. According to these authors, phenomenal consciousness consists of nothing more than such happy coalescence of enactive representations.

Problems with FAC-from-KAC Reductionism

It is difficult not to feel, however, that something is amiss in all of these accounts. There is a lingering impression that they presuppose that which they seek to explain and that without such illicit presuppositions the explanations simply do not work. Consider first Rosenthal's account. Sure, learning to make finer conceptual discriminations with respect to the taste of a wine, or to the sound of an oboe, allows for more refined experiences of the kind savoured by the connoisseur, but such finesse is the result of a process of training involving interaction between newly acquired knowledge and previous, more basic, experience. The connoisseur and the layperson experience taste, sound, or sight differently but they both operate within a space which is *already* richly experiential; the connoisseur is a specialist processor of experience, not the generator of experience out of insentience. Nevertheless, Rosenthal believes

[3]The idea that the explanation of phenomenal consciousness calls for a division of labor based on the conceptual distinction between a *qualitative* component (i.e., the something it is like aspect of seeing x) and a *subjective* component (the "for me" aspect of private experience) is due to Levine (2001). Kriegel (e.g., 2005, 2009) developed an influential self-representational account of consciousness in which this conceptual distinction plays a prominent role. However, unlike Janzen, Kriegel was never quite as adamant to declare that implicit self-awareness fully explains phenomenal character and recently he came to concede that a materialist reduction of phenomenal consciousness remains an elusive goal (Kriegel, 2011).

that the lesson from the connoisseur analogy is that there is a direct proportionality between the conceptually savvy and the experientially potent, a lesson which he then extrapolates to argue that peeling away all HOTs would eventually culminate in the absence of all phenomenal consciousness.

This suggestion, however, is highly non-intuitive. The idea that experience is constitutively dependent on conceptualization seems to put the cart before the horse: it would make the capacity to sense and feel which, by all accounts, is rather basic in evolutionary terms dependent on the abstract operations of conceptual thought — a much more sophisticated agent, and an evolutionary latecomer. Nor does the suggestion seem to fit with phenomenological data. Psychoactively induced experiences are, at times, remarkably rich, despite, and perhaps partly because of, the fact that the veil of rational classification and control is being lifted. Even more so, mystics of all ages consistently report that the cessation of all thought through meditation leads not to numbness and stupor but to the most intense and lofty experiences possible for humans (or think of the vividness, intensity, and freshness of the experiential reality of a child, which stands in sharp contrast to the child's lack of conceptual sophistication).

Viewed from a different angle, we may question not only the plausibility of Rosenthal's proposal but also its very intelligibility (cf. Goldman, 1993). For how could the mere fact that M is being represented by a higher-order thought M* turn M into a phenomenally conscious state? Crucially, there is nothing in this scenario which implies an internal modification of M, let alone a radical modification of the sort which would be required in order to make it a locus of sentience. The relation of being represented by M* is, insofar as M is concerned, an external relation, which implies, in turn, that whether or not the relation holds is something which has no effect, or need have no effect, on M's intrinsic qualities (compare: the fact that I represent the Eiffel tower as I think of it now induces no visible change in the tower itself, something more needs to be added if such a change is to be effected). How, then, could such a relation turn M from an insentient state (as per hypothesis) to a vehicle of sentience? The transition seems miraculous enough even if M* did have a clear causal impact upon M, let alone when it has none![4]

A careful reading of Rosenthal reveals that he also does not believe in such alchemy, arguing that being transitively conscious (viz., aware) of a sensory state M does not change the properties of that state. Rather, the effect of the conceptual HOTs we apply to M is to "enable us to be conscious of sensory qualities we already had, but had not been conscious of" (2002, p. 413). But the problem refuses to go away: the idea that awareness of M's sensory qualities (courtesy

[4]It may be mentioned in passing that Rosenthal's conviction that phenomenal consciousness can be explained in strictly extrinsic terms is not shared by all HOT theorists. Gennaro (1996), for example, developed a HOT account which strives to accommodate the intrinsic character of non-transitive conscious mental states.

of the HOT M*) is the key for explaining sentience is illusory. For if we assume that both M and M* are wholly insentient then neither the qualities of M, nor the medium whereby they are represented in M*, are truly experiential. At best, all there is here is information in one wholly insentient physical state M* about the properties of another wholly insentient physical entity M (and note that the fact that the relevant properties are *sensory* properties makes no difference: for unless we assume a reading of "sensory properties" which illicitly introduces the reality of sentience then such properties would be just like any other properties of numb matter). In short, it is hard to see where in all of this phenomenal consciousness could ever be found. To suppose that informational liaisons between wholly insentient internal states could somehow result in there being sentience somewhere in the system is to expect the blind to successfully lead the blind, but, as the old saying goes, "when the blind lead the blind both shall fall into the ditch" (Matthew 15:14).

Nor do I think that the other theories mentioned above fare better. Janzen's attempt to derive phenomenal character from implicit self-awareness ultimately faces the same problem. Self-representation is an internal relation, which is to say that the act of representation whereby P (a perceptual state) represents itself is part and parcel of P's identity. This means that the "no difference" argument raised against Rosenthal cannot be raised here because self-representation does make a difference — P would not be quite the same if it was not self-representing. However, the problem is with the suggestion that the difference which this relation of self-representation brings about, or explains, is the difference between sentience and utter insentience. That is, if we assume a basic ontology of insentient matter, and if we assume that self-representation is the one crucial factor which delivers us onto the realm of sentience, we end up with absurdity.

The absurdity lies in the idea that the fact that a given state M curls upon itself, thereby instantiating an informational closed-loop, could somehow turn it into a locus of experience. For, how could M's self-accessing be responsible for the transmutation? If the default assumption is that barring the self-referencing, M is just like any other physical state, which, per hypothesis, means an utterly insentient state, then there is "no one at home" to feel, sense, or be cognizant of the incoming information, and the fact that the information is self-originated, or self-effected, does nothing to change it. Or to put it differently, an insentient medium, or substance, cannot feel itself any more than it can feel any other thing — for it can feel nothing at all. Figuratively speaking, to suppose that self-representation can generate sentience from scratch is like supposing that a blind person can gain sight by staring at her own reflection in the mirror (see Levine, 2006, for an alternative argument against the idea that self-representation holds the key for solving the hard problem of consciousness).

The problem recurs in a different guise for theories that emphasize the role of the thick present moment. The coalescence of temporally differentiated rep-

resentations onto a unified thick moment may constitute a significant step in solving the problem of explaining the possibility of a temporally extended awareness, that is, of an awareness which goes beyond the instantaneous moment of physical time, but it does nothing to usher in phenomenal consciousness. The contribution of the thick specious present is purely functional: it consists of enabling us to dwell on our experiences and to savour them, granting us access to an inner reality which would otherwise go unnoticed (see Ellis and Newton, 2005). But note that this could only work if we sneak in the implicit assumption that the reality which is thereby accessed is experiential through and through — clearly an illicit move for anyone who proclaims to explain the coming into being of phenomenal consciousness. For if we assume, as an orthodox physicalist should, that the relevant representations, call them M_P, M_N, and M_F (for past, present, and future), are decisively insentient, then it makes no sense at all to suppose that the operation of fastening them together such that there is a partial overlap between them could somehow result in there being a vividly phenomenal character to their overlap. Rather, we should expect to get just what we were constructing: a superposition of phenomenally vacuous states.

The moral of these consistent failures is that you cannot derive experience from information processing; you cannot get FAC from KAC. Representational access may serve to *transform* phenomenal character in myriad significant ways but it cannot generate sentience from scratch. If we start with the idea that the task is to derive experience from a decisively non-experiential realm, and that representations of one sort or another are our means to do so, we just end up producing more and more blind representations, blindly representing a blind world. Or in Levine's apt words: "it's just piling on more representations" (2006, p. 195).

If the FAC-from-KAC hypothesis is a dead-end street then the physicalist who is also a qualia realist and qualia internalist has to concede that the emergence of sentience out of an insentient world cannot be explained in terms of intra-representational accessibility. The alternative which seems to force itself upon her is that such emergence is a pure physical fact, devoid of an epistemic rationale. Somehow, certain complexly interacting organizations of matter, in particular neural activation patterns, manage to evoke experience as part of their activation, but we cannot explain such evocation in psychological terms.

This concession is somewhat disturbing given the ample evidence which suggests that sentience and sapience (FAC and KAC) go hand in hand and are intimately connected, but it may not be so worrisome if other sciences could step in and fill the explanatory gap. Yet, this hope, too, seems to be in vain. As Nagel (1974), Levine (1983), Chalmers (1995), and others have pointed out, the emergence of sentience against the background picture presupposed by orthodox physicalism

has the appearance of a hopelessly brute fact, and a high-level brute fact at that.[5] It is no more explicable in physical, chemical, biological, or computational terms than it is in strictly psychological terms. The nagging question, "Why sentience?" remains just as vexing, no matter which scientific discipline we care to consult.

Indeed, the conceptual aporia to which the FAC-from-KAC hypothesis leads are but special exemplifications of the more general problem known as the explanatory gap (Levine, 1983), or the hard problem of consciousness (Chalmers, 1995). If this is the case, one may wonder why I have bothered to discuss at length the theories mentioned above only to end up returning to such a familiar point. The answer is that a firm understanding of the reasons behind the failure of the FAC-from-KAC hypothesis is necessary in order to motivate an alternative, non-reductive approach towards understanding the relations between sentience and sapience. In the remaining sections I present and defend the essentials of such an alternative.

Moderate Emergence and the Continuity Principle

As an alternative to the idea that sentience is derivable from epistemic liaisons between insensate mental states, I shall now pursue the idea that sentience and sapience go hand in hand — both co-evolve in correlation and none is more fundamental than the other. This parallelism is essentially in line with Chalmers' (1996) *coherence principle*. However, Chalmers' view has certain concomitant components which I am reluctant to accept, in particular: (a) his functionalist principle of organizational invariance according to which all functional isomorphs of a given conscious system S are experientially indiscernible from it; and (b) his analysis of awareness in terms of information processing, understood in the strictly syntactic sense of information theory. None of these latter elements is included in the present account.

Be that as it may, there is a more general issue which we must address before returning to the specifics of my proposal, to wit: What are the natural boundaries within which we should expect to find consciousness, complete with its correlative knowledge aspects and feeling aspects? Logically speaking, a parallelism between sentience and sapience implies only that they come together, but it tells us

[5]As Levine (1983) points out, certain physical facts simply *are* brute facts (at least from our human limited perspective). Yet, as he observes, the arbitrariness of sentience is particularly disturbing precisely because, unlike other brute facts such as the particular value of the gravitational constant, sentience is presumed to be a higher-level phenomenon, and higher-level phenomena are alleged to be *explicable* in terms of our theoretical understanding of the workings of lower-level phenomena. Chalmers' plea for an ontologically *fundamental* theory of consciousness (1996, chap.8) can be seen as an attempt to correct this anomaly by situating psychophysical laws alongside the basic laws of physics.

nothing about their scope in the natural world. In particular, parallelism does not tell us whether consciousness is ontologically primordial or whether it is an evolutionary latecomer. However, given the earlier observation that the orthodox physicalist picture seems to render the reality of conscious experience irredeemably inexplicable, it stands to reason that if we wish to avoid this most unpleasant consequence we ought to open ourselves to the possibility that nature's default state may not be categorically exclusive of sentience and subjectivity. And since orthodox physicalism is implicitly committed to the idea that consciousness, if it exists at all, is a radical ontological emergent, a consistent alternative is likely to involve the idea that consciousness is a moderate emergent.

Moderate emergence is the thesis that there is continuity in cosmic evolution, such that if X emerges from background conditions $C_1 \ldots C_n$ there must be something about $C_1 \ldots C_n$ which, in principle, could render X's emergence, and its unique characteristics, intelligible. In other words, the seeds of that which emerges must somehow be latent already within that from which it emerges (for historical precursors to this idea see Leibniz, 1704/1995a; and Peirce, 1892/1955; in particular the former's *law of continuity*, and the latter's concept of *synechism*). A special corollary of moderate emergence is that the emergence of creatures endowed with an internal dimension out of physical preconditions in which such a dimension is presumed completely absent is precluded on pain of violating the continuity principle. Thus, on the assumption that an intrinsic dimension is clearly manifest in the structure of our own consciousness, it follows that such a dimension must be an integral part of nature at all levels of organization.

Those who espouse this line of reasoning often make use of what Seager (2006) calls the *intrinsic nature argument*. Following the footsteps of Eddington (1928) and Russell (1927), they note that scientific explanations are limited to the structural–dispositional aspects of reality, leaving unaccounted the intrinsic nature of the entities which science purports to describe and explain. Thus, it is not so much that modern science denies the existence of such intrinsic natures, or qualities (let alone proves their inexistence) but, rather, that it ignores them. Combined with the claim that intrinsic qualities are a logical desideratum, and that consciousness provides us with an existential proof of their reality, it is then suggested that there is more to reality than what is currently subsumed under the conceptual umbrella of contemporary natural science, and that a more complete metaphysics will have to take into account the intrinsic nature of things (advocates of this line of reasoning include Chalmers, 1996; de Chardin, 1959; de Quincey, 2002; Lockwood, 1989; Maxwell, 1979; Nagel, 1979; Rosenberg, 2004; Seager, 2006; Shimony, 1997; Stoljar, 2001; Strawson, 2006).

The significance of the intrinsic nature argument in the current context lies in the fact that the argument provides elbow room for the scenario of moderate emergence. In a world where nature, in its primordial state, lacks intrinsic qualities, the emergence of sentience is destined to constitute a radical and

inexplicable ontological breach, but in a world where such qualities exist, and are the rule rather than the exception, the emergence of macro phenomenal consciousness may simply represent a natural outgrowth out of humbler origins of a similar kind.

However, there is much disagreement over the question how to interpret this kinship and the ontological continuity it implies. In particular, does having an intrinsic nature imply having sentience; or does it merely imply a certain potentiation towards sentience. Panpsychists take the continuity principle to imply that sentience scales all the way down, or, in other words, that experiencing subjects and their corresponding phenomenal properties are aboriginal. To be sure, the experiential reality of an atom, an organic molecule, a metazoan, or a primitive protozoan is very different from ours, but, the idea goes, they nevertheless enjoy certain experiences (present day defenders of panpsychism include, for example, de Quincey, 2002; Griffin, 1998; Rosenberg, 2004; Seager, 2006; Sprigge, 1983; Strawson, 2006). In contrast, others, whom we may identify as Russellian identity theorists, or panprotopsychists (see Chalmers, 2013), argue that the intrinsic natures of sufficiently primitive beings are wholly insentient and yet that such intrinsic natures are *proto*-phenomenal in the sense that, when properly combined, they instantiate experience in an intelligible manner (defenders of this view include, for example, Feigl, 1958; Lockwood, 1989; Maxwell, 1979; Pereboom 2011; Stoljar 2001). Finally, there are also those who endorse neutral monism in the tradition of Mach (1886/1959) and James (1912) and argue that the fundamental entities are phenomenal properties but that subjects capable of experiencing such properties emerge only at a later stage (for a recent defence of this view see Coleman, 2014).

The theoretical framework I shall present shortly is panpsychist. On this occasion, I make no systematic attempt to motivate panpsychism over and against the other positions just mentioned. Nor do I offer a defence of panpsychism against the charge that it faces a combination problem (the term is Seager's, 1995) which is every bit as hopeless as the hard problem of consciousness that haunts orthodox physicalism. Doubtless, these are issues which sympathizers of panpsychism must address and I have done, to some extent, elsewhere (Shani, 2010). However, on the present occasion my goal is not to validate panpsychism fair and square but, rather, to explore the modifications which such a view entails with regard to the relevance of awareness for the explanation of sentience. Ultimately, I argue that a panpsychist framework provides a more coherent picture of this explanatory relation than the one bequeathed upon us by physicalism. If my diagnosis is correct, then it ought to serve as yet another reason to resist orthodox physicalism while moving in the direction of assigning consciousness a greater role in the scheme of things. However, I must qualify myself by adding that even if my point is valid we cannot rule out, at this stage, the possibility that certain alternative monistic positions other than panpsychism — perhaps

neutral monism, or perhaps panprotopsychism — might be able to claim an equal degree of explanatory coherence with respect to the problem at hand.

More FAC through More KAC: Outlines of an Ampliative Approach

Having rejected the idea that epistemic liaisons are constitutive of phenomenal consciousness I propose instead that the contribution of sapience for the making of sentience is ampliative. On the ampliative model, informational liaisons modulate phenomenal character rather than generating it from scratch. Moreover, the modulation is enhancive, which is to say that the there is a positive correlation between the representational complexity and power of a system and its phenomenological richness — an increase in one is conducive to an increase in the other. Correspondingly, from this perspective, the main explanatory challenge does not consist in explaining how phenomenal consciousness comes into being in the first place but, rather, in explaining how it *changes* as a function of changes in representational power.

The ampliative model can be characterized by five basic theses:

1. [Concomitance]: Every act of presentation, or of re-presentation, involves a subject in cognizance of an object, or a datum, which it presents, or represents, through a subjective medium, which reacts to the object, or datum, with feelings.[6]
2. [Endo-phenomenology]: Phenomenal character is an endogenous feature of the medium of representation; which is to say that even in the absence of stimulation the medium is still a locus of sentience.
3. [Transformation]: Presentational, or representational, acts operate on the medium as transformative agents, constraining and modulating the ever-present flow of experience.
4. [Correlation]: In general, there is a direct proportionality between the level of sophistication of a system's cognitive organization and the depth and variability of its phenomenal world.
5. [Enhancement]: Informational liaisons between representational states are often instrumental in enriching the structure and character of experience. In other words, the transformative effect of acts of awareness on the subjective medium is often in a qualitatively ascending direction.

[Concomitance] is reminiscent of Whitehead's (1929/1985) notion of *prehension*. The important point in the present context, however, is the concourse between

[6]The distinction between *presentation* and *representation* parallels Searle's (1983), which means that it corresponds to the distinction between those situations in which the intentional object is present to one's senses and those in which the intentional object is not currently present and has to be represented in one's mind. For simplicity's sake, however, I will follow the common practice of using the term "representation" in a looser sense covering both presentations and re-presentations.

an endogeneously sentient substrate (viz., the medium) and the acts of awareness, or sensitivity, whereby the subject takes into account external data. In other words, the idea is that sentience and sapience are coextensive: a sentient system is simultaneously a system which exemplifies awareness to events within, and outside, itself; likewise, a system capable of genuine awareness is, concurrently, a sentient system.

[Endo-phenomenology] expresses the idea that the physical substrate which serves as a medium for occurrent conscious representations is inherently sentient. This means that in the absence of significant external stimulation such a medium maintains a relatively homogenous qualitative state (in a manner analogous to that of an energy field subject to no discernable local excitation), or, alternatively, that it generates its own activation patterns, perhaps subject to chance events. External stimuli create stirs, or waves, on the surface of this "ocean" of spontaneous activity which in turn effect further transformations down the line, inducing changes in the patterns of organization that characterize the medium at the time. Thus, the structure of the medium (of which phenomenal tone is an essential aspect) is responsive to the structure of the environments with which the subject interacts.

[Transformation] serves to emphasize that the ever changing flow of experience is modulated by representations. That is, both representations of the outside world (by way of anticipation, perception, memory, or imagination) and of the self (i.e., representations of activities within the system or of the manner in which the system is influenced by external encounters) induce changes in the structure and course of the system's internal experiential flow, leading to consequent representations and consequent process modulations down the line.

[Correlation] stresses a direct proportionality between the representational complexity exemplified by a cognitive agent and the phenomenal riches which the agent enjoys (or can enjoy). To use an extreme example, there is little reason to doubt that the mental reality of an orangutan is considerably richer than that of a jelly fish, not only on account of cognitive sophistication but also in terms of phenomenal variability and depth. These differences are indicative of a general rule, applicable throughout the animate world: the higher we go up the evolutionary ladder we find greater riches both in terms of sapience and in terms of sentience (at the same time, we must guard against the tendency to downplay the emotional and cognitive sophistication of relatively simple creatures, or to ignore their uniqueness). Conversely, if experience is something which even inanimate entities are presumed to possess, it is natural to expect this general rule to continue to hold all the way down so that, in the words of Teilhard de Chardin, "*[r]efracted rearwards along the course of [cosmic] evolution, consciousness displays itself qualitatively as a spectrum of shifting hints whose lower terms are lost in the night*" (1959, p. 59, italics in the original).

Finally, [Enhancement] drives home the point that the more a system is capable of accessing and processing its own experiences the more it is capable of having experiences of novel kinds, thereby intensifying its experiential reality. Consequently, [enhancement] constitutes the one aspect of the ampliative model most relevant for the present discussion, and the one on which I focus henceforth.

That [Correlation] obtains is something which calls for an explanation, and the best explanation seems to be that our two complementary dimensions of conscious experience — sentience and sapience — are mutually reinforcing. The gist of the idea is that greater representational power is conducive to greater variability and intensity in a creature's phenomenal life; and collaterally, an increase in the scope and intensity of phenomenal expression augments the capacity for representational classification, leading to novel and more articulated forms of awareness, and of action-guidance through awareness. Mutual reinforcement is, of course, a bidirectional relation but in line with my earlier resolve I focus here on the amplificatory effect of incremental awareness on phenomenal consciousness.

For illustrative purposes, imagine a creature which we may call Primo. Primo is a blobby little creature whose protoplasmic interior manifests a minimal degree of internal organization. It is, however, sentient. It detects certain chemical gradients, and reacts to light, heat, and mechanical contact. These environmental interactions translate to internal events one aspect of which is that they create ripples in Primo's drearily shallow endo-phenomenological pond. Some of these ripples are recurrent and systematic enough to play a role in guiding Primo's behaviour. For the most part, however, ripples (whether spontaneous or externally induced) appear across the pond only to disappear quickly without leaving visibly recognizable traces. Yet, Primo is a special creature. It goes through a developmental catastrophe after which it changes quickly and radically. It grows in size; its internal milieu differentiates to various compartments, giving rise to a multitude of well-coordinated organelles, cells, tissues, and organs; it also grows external organs, some specialized for locomotion and object manipulation, some for the detection of information; it even grows an impressively dense ganglia full of interconnected nerve cells which enable it to integrate information from its newly grown perceptual and motor organs (as well as bodily surface) with information from its newly grown internal milieu, to make records of such informational confluence, to recall traces of those records, and to use all of this in guiding the activities of its monstrously changed self. In short, Primo is a one-in-all evolutionary freak.

Clearly, we should expect post-catastrophic Primo, call it Primo_2, to enjoy a richer phenomenal reality than its pre-catastrophic self Primo_1, but the question is why. One explanation, which is in line with much of contemporary thinking

about panpsychism, is that this has to do with the fact that $Primo_1$ is but a tiny micro-organism whose phenomenal field is limited to micro-experiences with micro-phenomenal properties, whereas $Primo_2$ is a multi-cellular organism whose phenomenal field combines the phenomenal fields of its micro-components, giving rise to a rich tapestry of macro-experiences endowed with macro-phenomenal properties. Now, whether or not such a combination story makes sense (recall the combination problem) there is no denial that bulk is a factor in the differences between $Primo_1$ and $Primo_2$. To use aquatic metaphors, if $Primo_1$'s endo-phenomenological space is a shallow pond then $Primo_2$'s is a vast ocean, and, as we know, it takes an ocean to manifest certain wave patterns.

But, of course, this is only part of the story. Patterns of ripples and waves (our analogy for experiences) depend on other factors: wind currents, the moon, volcanic activity, local movements of vessels, objects, and animals, the throwing of stones, even artificial wave generators. In the end, what matters are the *patterns* of disturbance generated and bulk is, at best, only a necessary condition for that. To go back to the thought experiment, the moral to take home is that if we wish to explain the spectacular differences between the phenomenal realities of $Primo_1$ and $Primo_2$ we must look for the formative agency, the "wave generator" responsible for creating such vast differences in the patterns of disturbance characteristic of the respective endo-phenomenological media of these creatures.

To continue this idea, I think that the fact that the experiential life of $Primo_2$ is so much richer than that of $Primo_1$ depends crucially on the enormous differences in their degrees of internal *organization*. $Primo_2$ is a complexly organized creature capable of constraining, directing, and regulating the flow of energy, and the distribution of work, throughout itself in multitudinous ways unavailable to $Primo_1$. This increased capacity for self-governance is, I suggest, the formative agency we need to look at.

Clearly, the development of more powerful representational capacities is an aspect of advanced self-governance. It enables improved process coordination, anticipation, action-selection, and much more (for further discussion of the connection between representation and self-governance see Bickhard, 2000; Clark, 1995; Collier and Hooker, 1999; Kauffman, 2000; Pezzulo, 2011; Shani, 2006). The ability to know more, with better resolution, in greater detail, and with greater depth and scope allows for the possibility of more refined self-governance and opens up new horizons for practicing novel forms of interaction and self-conduct. This much is evident, but the reason I mention it here is the *formative* influence on the qualities of experience. Unlike $Primo_1$, $Primo_2$ enjoys vast representational resources: a wide spectrum of sensory, somato-sensory, motor, and visceral differentiations, which enable the formation of a plurality of perceptual and other presentational states; the ability to memorize, and to re-enact memorized representations; a capacity to form prospective representations anticipating

future conditions; powerful means for processing and integrating cognitive information; the ability to monitor its own internal events and to interact with them, using higher-order states, etc.

My point is that these functional aspects serve to augment and enrich experience. $Primo_1$ has a very limited access to the world as well as to its own internal conditions. In contrast, $Primo_2$ has a broader, deeper, and more articulated access to the world around it, and in addition it also has far more sophisticated ways to access its inner reality. These windows on the world and on the self are really operations which induce novel and ever more refined disturbance patterns in the creature's endo-phenomenological space, thereby enriching the landscape, or texture, of that space.

It is easy to see how a greater ability for making perceptual discriminations augments one's phenomenal world — it creates more experiences, and more shades of experience. But so is the case with the capacity to operate upon one's own representations, to experience one's own experiences, as it were. When an image is recalled and re-lived in the light of present experience, when one's own feelings are addressed, when a connection between different elements in one's experience is discerned and illumined by awareness; in short, whenever consciousness loops upon itself and the flow of experience becomes an object of experience, new *types* of experience emerge which were not available before. This recursive process is seemingly boundless — there are always novel and more refined experiences to be distilled, provided that the distillery (viz., the system's organization) is up for the task.

Thus, the difference in Primo's experiential life before and after the morphogenetic mutation is, in large part, a difference in the capacity of its bodily organization to whip the waters of consciousness into shape. This, then, is the idea behind the ampliative model: that an increase in the capacity for information processing and access transforms and enriches the texture of one's phenomenal life without, however, being responsible for the fact that there is experience in the first place.

Putting the Ampliative Approach to Work

Above, I criticized theories committed to the FAC-from-KAC hypothesis for being caught in an explanatory cul-de-sac. I now proceed to show that once we translate these theories from their natural reductive setting to the non-reductive landscape delineated by the ampliative approach, we can restore coherence to some of their more attractive features — although, naturally, this process involves a reinterpretation of the meaning and scope of these theories.

Recall, first, Rosenthal's HOT-based account of phenomenal consciousness. I argued that the idea that a mental state M could become phenomenally conscious in virtue of being represented by a higher-order thought M* defies sense,

but now look at the situation from the perspective of the ampliative model. The model assumes that the higher-order monitoring process M*→M explains neither why M is phenomenally conscious, nor why M* is. However, it predicts that such higher-order monitoring will result, typically, in experiences which are novel in kind, i.e., experiences of a kind whose existence is contingent on this very process. But where should we look for such experiences? Clearly, the object-level, the level at which M itself is located, would be the wrong place to look for such emergent phenomenology since, as mentioned earlier, the mere fact that M* represents M does not imply any modifications in M itself (unless, of course, the higher-order monitoring process is an intervening one). Rather, it is to the meta-representational level, M*'s level, that we should turn.

At the meta-level (or levels), we find mental states whose intentional objects are other mental states, and which represent qualities of those object-level states even as the latter represent qualities of the environment, or of the body (Bickhard, 2005). Thus, the features represented at the meta-level are different than the ones represented at the object-level. In particular, they may include such abstract elements as relations among the contents of object-level mental states, relations such as causality, similarity, ordering, matching, etc. (see, for example Barsalou, 1999; Chapman and Agre, 1986; Pezzulo, 2011). And awareness of such relations (perhaps courtesy of levels of representation higher-up the hierarchy) engages novel experiences, including an experiential type which is crucial for the discussion below, namely, the experience of *feeling oneself as an integrated experiential subject*. Thus, there is a grain of truth in Rosenthal's claim that higher-order cognitive processes are conducive to more refined phenomenologies; it is just that we can't expect meta-cognition to be the ultimate explanation of the reality of phenomenal consciousness.

Nor can we expect same-level self-awareness to carry the task. Above, I argued that self-awareness is of little help as long as we continue to assume that that which is being accessed, or in this case that which accesses itself, is inherently insentient. For if a mental state M is realized in an utterly insentient medium then it can neither be a locus of experience, nor can reflexive access grant it acquaintance with its own (non-existent) "experiential content." However, as soon as we change our default axiom to one in which experience occupies a fundamental place in nature, things begin to make better sense. First, hypothesis M is now realized in a medium which is inherently sentient, hence we should have no problem understanding how it could be a locus of experiential content. Second, we can now begin to make sense of the import of self-awareness: for if M, an inherently sentient state, represents itself, we should expect such access to yield conscious awareness of the experiential content enfolded in M. Such awareness would take the form of experiential acquaintance with M's base-level experiential content, where both the base-level experiential content and the higher-level acquaintance with that content are

complementary aspects of M's phenomenal portrait. Supporters of the self-representational view are correct to emphasize the significance of self-awareness, for, clearly, the ability to be aware of one's inner reality is a crucial ingredient of consciousness *as we know it*, that is, as we find it in the structure of human phenomenology (see, for example, Zahavi, 2005 p. 24). Their mistake lies in the failure to realize that the contribution of self-awareness is intelligible only against a background which is already sentient.[7]

Lastly, consider the specious present account. The ampliative model is rather congenial to the idea that the ability to experience past, present, and future in a single "specious moment" is a significant landmark of consciousness. However, a careful scrutiny of this idea reveals that its real value lies not in the fact that it explains the transmutation of utterly insentient representations into a single complex locus of sentience, for this it does not do. Rather, the real contribution of the specious present with respect to phenomenal consciousness lies in the fact that it equips cognitive agents with a temporal window wide enough to enable us to become acquainted (i.e., experientially acquainted!) with our ongoing experiential flow.[8] As such, it constitutes a major step in the discovery that we are enduring subjects of experience, but it does not explain the emergence of experience from the non-experiential.

Time and again, then, we see that the real contribution of sapience to the explanation of phenomenal consciousness is transformative and ampliative: it is instrumental in explaining how novel qualitative types of experience emerge atop other, more basic ones. Yet, no matter how hard we search, never do we find a single instance in which epistemic liaisons generate sentience from scratch. More from less everywhere, but nowhere is there something from nothing.

The Discovery of Experience: A Layered View of the Evolution of Consciousness

While epistemic access does not, and cannot, beget phenomenal consciousness, it plays a crucial role in explaining an important stage in the evolution of conscious experience, namely, that stage wherein a system acquires the capacity to *experience itself as an integrated subject of experience*. In other words, there is, indeed, a sense in which sapience is indispensable for an explanation of sentience but

---

[7]It might be the case that some supporters of the self-representational view (especially within the phenomenological tradition) are not committed to the constitutive approach and may even be sympathetic to the point I am making, yet I'm unfamiliar with any clear admission of this point. Thus, whether the point is denied, or whether the issue is insufficiently clarified, the overall impression is that advocates of the self-representational view succumb to the FAC-from-KAC fallacy.

[8]This idea is stated rather clearly by Ellis and Newton (2005) except that they fail to see with sufficient clarity that, as a matter of fact, what they explain is not our capacity to experience in the first place but, rather, our capacity to experience our own experiences!

this sense is limited to a higher-order form of phenomenal consciousness, consisting of sustained experiential acquaintance with one's own experiential life. Thus, it is not experience as such which is contingent on robust epistemic liaisons between inner mental states, but, rather, something different and more intricate, namely, the subjective discovery of the fact that one is the owner of an inner experiential realm! Such a sense of ownership over one's experiential domain ought not to be confused with full-blown *self*-consciousness since, in its most basic form, it implies neither possession of the concept of self, nor of a temporally extended ("autobiographical") sense of self (see below). Nevertheless, the capacity to sense the flow of one's experience as an integrated subjective arena is a precursor of mature self-consciousness, as well as of other high-level manifestations of reflective consciousness.

Prima facie, the idea that acquaintance with one's own experiential reality is an emergent phenomenon is provocative and even paradoxical since it can be easily interpreted as suggesting that below that level of emergence are creatures (or entities) which, although phenomenally conscious, are completely unaware of their inner realities. Now, this is a strange proposition. It is widely held that the very condition of being in a phenomenally conscious state implies awareness of that state (see, e.g., Chalmers, 1996; Kriegel, 2005). If so, then there can be no such thing as a phenomenally conscious yet wholly unnoticed, or unannounced, mental state, and this, in turn, cuts against the idea that it is possible for a creature to be phenomenally conscious without being in the least aware of its inner experiential flow.

In response, I should stress that I do not claim that it is possible to be phenomenally conscious without exemplifying *any* degree of awareness whatsoever. Nor does such a result follow from my analysis. Rather, my claim is more qualified, namely, that a certain degree of (emergent) epistemic access is a prerequisite for a certain degree of reflexive acquaintance with one's experiential flow, and that to the extent that such a degree of epistemic access is compromised it also compromises one's familiarity with one's underlying phenomenal reality. Or to put it in more concrete terms, the point I am making about higher-order experiential access to one's own experiences is that such access is a prerequisite for a *stable integrated acquaintance with one's inner reality, experienced as one's own* — I make no claim to the effect that a system which lacks such higher-order access to its own experiences lacks any kind of sensitivity whatsoever to its inner reality.

To illustrate the idea think first of a creature who is as simple as Primo_1, or perhaps even simpler. As mentioned before, such a creature would be subject to various kinds of experiences, various kinds of disturbances to the endogenous oscillatory patterns of its endo-phenomenology. Some disturbances would be powerful, systemic, or significant enough to consume the creature's attention (however diffusive or automated it may be), or to trigger adaptive responses.

However, per hypothesis, such a creature is extremely primitive: its capacity to retain traces of its past experiences and to re-enact such memories, as well as its capacity to form anticipatory images of future events, are so rudimentary that it is virtually confined to an eternal phenomenal present. Moreover, and crucially for our present concern, the creature lacks the ability to form higher-order representations which would enable it to reflect back on its subjective experiential flow, to integrate its base-level experiences into a meaningful (and accessible) whole, and to appreciate the qualities and interrelations of such experiences. Thus, it has no means of discerning lasting relationships between classes of internal events, and of consciously recognizing their significance. In short, the poor creature's phenomenal world is both punctated and flat: fleeting experiences come and go like actors on stage but the agent having those experiences is just too amorphous to maintain a clear sense of ownership over the show.

Admittedly, this scenario is somewhat extreme, but we need not be afraid to think in extreme terms when probing into the possibility of sentience below the level of multi-cellular organisms, let alone below the bar of biological existence. The point I wish to stress is that a hypothetical creature of the sort just imagined illustrates the possibility of having a phenomenal life while, at the same time, being almost totally unaware of the fact that one has such a life. Knowledge of the fact that one is the owner of a private domain, acquaintance with the secret of one's own subjectivity, requires much more. It is like a mystery into which only some are initiated (and then, only partially and gradually) namely, those creatures whose organizational features enable them to loop over themselves, turning their experiential flow into an object of experience and observing the privacy of their inner world unfolds.

A prominent contemporary advocate of the idea that a sense of ownership over one's subjective reality is an emergent construction contingent on higher-order modes of access is Antonio Damasio (1999). Damasio's view of consciousness and selfhood is a layered view in which cognition builds upon emotion, which in turn builds upon homeostatic regulation (see also Dempsey and Shani, 2013; Watt, 2004). At the basis of his analysis is a layer he calls the *proto-self*, which is "a coherent collection of neural patterns which map, moment by moment, the state of the physical structure of the organism in its many dimensions" (p. 154). As the organism interacts with items in its environment it forms images of these items, and the proto-self undergoes modifications in response to these images, which, in turn, give rise to emotional reactions, and to mental images, or "feelings," responsive to such reactions. Now, according to Damasio, these patterns of interrelationships between images of the items with which the system interacts and the corresponding modifications to the system's proto-self are captured by higher-order representations recording the manner of change. In turn, these higher-order representations are responsible for the construction of a higher-layer of selfhood which Damasio calls *core consciousness*, and that

consists of a momentary sense of self in the act of being internally modified. Core consciousness, Damasio argues, provides for a sense of ownership over one's own inner reality, albeit a rather basic sense which is neither temporally extended (autobiographical), nor one which depends on verbal ability or on moral sense. It is a sense of ownership over one's private world which, presumably, many animals are capable of exemplifying but which requires a level of sophistication far beyond that of a creature like $Primo_1$.

What makes Damasio's account apt for the present discussion is the fact that it provides an empirical model (which incidentally also hints at the significance) of the ontological partition between (i) being an abode of subjective experience; and (ii) being capable of experiencing one's own subjective domain *as one's own*. Unfortunately, Damasio is somewhat unclear as to whether "the feeling of what happens" effected by his higher-order representations is a feeling of lower-order experiences or, rather, of utterly insentient occurrences. While only the first interpretation corresponds to the view I advance here, I believe that his work helps clarify the claim that an ontological partition of the sort just mentioned is a potentially important one.[9]

Such a partition was emphasized by Leibniz in his distinction between perception and apperception — the former being an inner state of the monad representing external things, while the latter consists of reflective knowledge of this inner state and is the prerogative of true minds (see Leibniz, 1714/1995b, 1714/1995c). It is also echoed in Whitehead's distinction between prehension and consciousness and in his claim that "consciousness presupposes experience" (1929/1985, p. 53). Leibniz complained that this failure to appreciate that not all perceptions are apperceived, or to put it in our terms, not all states of experience are objects of experience, led the Cartesians to their notorious belief "that [rational] minds alone are monads, and that there are no souls in animals, and still less other *principles of life*" (1714/1995b, p.197).

Few of us today would adhere to such stark Cartesianism yet, clearly, our collective legacy is much more Cartesian than Leibnizian. Doubtlessly, this legacy has some role to play in the reluctance of many to give any credibility to the possibility of hidden grades of consciousness scaling down throughout the whole of nature. Moreover, the fact that reflective awareness plays such a prominent

[9]On the one hand, Damasio describes his higher-order representations as "feelings of feelings" in a way which suggests that his higher-order representations are higher-order experiences representing lower-order experiences. On the other hand, since core consciousness materializes only at the level of higher-order representation his account leaves lower-order "feelings" below the bar of consciousness. It is interesting to note, however, that Damasio's theory of core consciousness is not aimed at explaining phenomenal consciousness as such but, rather, the sense of familiarity, identification, and ownership, which one feels with respect to one's inner reality (hence, the *scire* or knowledge connotation of "consciousness"). Thus, I believe that in essence his theory is consistent with an interpretation according to which the "feeling of what happens" is the sensing of one's own experiential flow.

role in the making of human consciousness makes it doubly difficult for us to think or to imagine clearly the possibility of conscious forms lacking all but the most primitive forms of awareness, or, to put it in Leibniz' terms, to conceive of monads which are not minds, or rational souls. Leibniz's warning against the potential detrimental consequences of failing to see behind the veil of our own phenomenology remains as relevant today as it was in his day.

References

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–600.

Bickhard, M.H. (2000). Autonomy, function and representation. *Communication and cognition — artificial intelligence* [Special issue on: The contribution of artificial life and the sciences of complexity to the understanding of autonomous systems], *17*(3–4), 111–131.

Bickhard, M.H. (2005). Consciousness and reflective consciousness. *Philosophical Psychology, 18*, 205–218.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18,* 227–247.

Brentano, F. (1995). *Psychology from an empirical standpoint*. London: Routledge. (originally published 1874)

Carruthers, P. (1996). *Language, thought and consciousness*. Cambridge: Cambridge University Press.

Chalmers, D.J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2,* 200–219.

Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory.* New York: Oxford University Press.

Chalmers, D. (2013). *Panpsychism and panprotopsychism*. The Amherst Lecture in Philosophy 8: 1–35. http://www.amherstlecture.org/chalmers2013/

Chapman, D., and Agre, P. (1986). Abstract reasoning as emergent from concrete activity. In M. P. Georgeff and A.L. Lansky (Eds.), *Reasoning about actions and plans — Proceedings of the 1986 workshop* (pp. 411–424). San Mateo, California: Morgan Kaufmann.

Clark, A. (1995). Moving minds: Situating content in the service of real-time success. *Philosophical Perspectives, 9,* 89–104.

Coleman, S. (2014). The real combination problem: Panpsychism, micro-subjects, and emergence. *Erkenntnis, 79,* 19–44.

Collier, J.D., and Hooker, C.A. (1999). Complexly organized dynamical systems. *Open Systems and Information Dynamics, 6,* 241–302.

Crick, F.H.C., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2,* 263–275.

Damasio, A. (1999). *The feeling of what happens*. New York: Harcourt, Brace, and Co.

Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79,* 1–37.

Dempsey L.P., and Shani, I. (2013). Stressing the flesh: In defence of strong embodied cognition. *Philosophy and Phenomenological Research, LXXXVI,* 590–617.

de Quincey, C. (2002). *Radical nature: Discovering the soul of matter*. Montpelier, Vermont: Invisible Cities Press.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge, Massachusetts: MIT Press.

Eddington, A. (1928). *The nature of the physical world*. New York: MacMillan.

Edelman, G. (1992). *Bright air, brilliant fire.* New York: Basic Books.

Ellis, R.D., and Newton, N. (2005). The unity of consciousness: An enactivist approach. *Journal of Mind and Behavior, 26,* 225–280.

Feigl, H. (1958). The "mental" and the "physical." In H. Feigl, M. Scriven, and G. Maxwell (Eds.), *Concepts, theories, and the mind body problem*. Minnesota Studies in the Philosophy of Science, Volume 2. Minneapolis: University of Minnesota Press.

Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge, Massachusetts: MIT Press.

Gennaro, R. (1996). *Consciousness and self-consciousness: A defense of the higher-order thought theory of consciousness*. Amsterdam: John Benjamins.
Goldman, A. (1993). Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition, 2,* 364–382.
Griffin, D.R. (1998). *Unsnarling the world-knot: Consciousness, freedom, and the mind–body problem.* Eugene, Oregon: Wipf and Stock.
Humphrey, N. (1992). *A history of mind: Evolution and the birth of consciousness.* New York: Simon and Schuster.
Humphrey, N. (2000). How to solve the mind–body problem. *Journal of Consciousness Studies, 7,* 5–20.
James, W. (1912). *Essays in radical empiricism*. New York: Dover.
Janzen, G. (2006). Phenomenal character as implicit self-awareness. *Journal of Consciousness Studies, 30,* 44–73.
Kauffman, S.A. (2000). *Investigations*. New York: Oxford University Press.
Kriegel, U. (2005). Naturalizing subjective character. *Philosophy and Phenomenological Research, 71,* 23–57.
Kriegel, U. (2007). Philosophical theories of consciousness: Contemporary western perspectives. In M. Moscovitch, E. Thompson, and P.D. Zelazo (Eds.), *Cambridge handbook of consciousness* (pp. 35–66). Cambridge: Cambridge University Press.
Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford: Oxford University Press.
Kriegel, U. (2011). Self-representationalism and the explanatory gap. In J. Liu and J. Perry (Eds.), *Consciousness and the self: New essays* (pp. 51–75). Cambridge: Cambridge University Press.
Leibniz, G.W. (1995a). New essays on the human understanding. In G.H.R. Parkinson (Ed.), *Philosophical writings* (pp. 148–171). (original work published 1704)
Leibniz, G.W. (1995b). Monadology. In G.H.R. Parkinson (Ed.), *Philosophical writings* (pp. 179–194). (original work published 1714)
Leibniz. G.W. (1995c). Principles of nature and grace. In G.H.R. Parkinson (Ed.), *Philosophical writings* (pp. 195–204). (original work published 1714)
Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64,* 354–361.
Levine, J. (2001). *Purple haze: The puzzle of consciousness*. New York: Oxford University Press.
Levine. J. (2006). Conscious awareness and (self) representation. In U. Kriegel and K. Williford (Eds.), *Self-representational approaches to consciousness* (pp. 173–198). Cambridge, Massachusetts: MIT Press.
Lockwood, M. (1989). *Mind, brain, and the quantum: The compound 'I.'* Cambridge, Massachusetts: Basil Blackwell.
Luu, P., and Tucker, D.M. (2004). Self-regulation by the medial frontal cortex: Limbic representation of motive set points. In M., Beauregard (Ed.), *Consciousness, emotional self-regulation, and the brain* (pp. 123–161). Amsterdam: John Benjamins.
Lycan, W. (1996). *Consciousness and experience.* Cambridge, Massachusetts: MIT Press.
Mach, E. (1959). *The analysis of sensations and the relation of the physical to the psychical [Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen;* C.M. Williams, Trans.]. New York: Dover. (Originally published 1886)
Maxwell, G. (1979). Rigid designators and mind–brain identity. In C.W. Savage (Ed.), *Minnesota studies in philosophy of science* (Volume 9, pp. 365–403). Minneapolis: University of Minnesota Press.
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83,* 435–450.
Nagel, T. (1979). Panpsychism. In T. Nagel, *Mortal questions* (pp. 181–195). Cambridge: Cambridge University Press.
Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition, 14,* 30–80.
Peirce, C.S. (1955). The law of mind. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 339–353). New York: Dover. (originally published 1892)
Pereboom, D. (2011). *Consciousness and the prospects of physicalism.* New York: Oxford University Press.

Pezzulo, G. (2011). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind and Language, 26,* 78–114.
Rosenberg, G. (2004). *A place for consciousness: Probing the deep structure of the natural world*. New York: Oxford University Press.
Rosenthal, D.M. (2002). Explaining consciousness. In D.J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 406–421). New York: Oxford University Press.
Russell, B. (1927). *The analysis of matter*. London: Kegan Paul.
Seager, W. (1995). Consciousness, information, and panpsychism. *Journal of Consciousness Studies, 2,* 272–288.
Seager, W. (2006). The intrinsic nature argument for panpsychism. *Journal of Consciousness Studies, 13,* 129–145.
Seager W., and Allen–Hermanson, S. (2013). Panpsychism. In E.N. Zalta (Ed.), *Stanford encyclopaedia of philosophy*. http://plato.stanford.edu/archives/fall2013/entries/panpsychism/
Searle, J.R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge, Massachusetts: MIT Press.
Shani, I. (2006). Narcissistic sensations and intentional directedness: How second-order cybernetics helps dissolve the tension between the egocentric character of sensory information and the (seemingly) world-centered character of cognitive representation. *Cybernetics and Human Knowing, 13,* 87–110.
Shani, I. (2010). Mind stuffed with red herrings: Why William James' critique of the mind-stuff theory does not substantiate a combination problem for panpsychism. *Acta Analytica, 25,* 413–434.
Shimony, A. (1997). On mentality, quantum mechanics and the actualization of potentialities. In R. Penrose (with A. Shimony, N. Cartwright, and S. Hawking), *The large, the small, and the human mind* (pp. 144–159). New York: Cambridge University Press.
Skrbina, D. (2006). Beyond Descartes: Panpsychism revisited. *Axiomathes, 16,* 387–423.
Sprigge, T.L.S. (1983). *The vindication of absolute idealism*. Edinburgh: Edinburgh University Press.
Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research, 62,* 253–281.
Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies, 13,* 3–31.
Teilhard de Chardin, P. (1959). *The phenomenon of men*. London: Collins.
Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind.* Cambridge, Massachusetts: MIT Press.
van Gulick, R. (2001). Reduction, emergence, and other recent options on the mind–body problem. *Journal of Consciousness Studies, 8,* 1–34.
van Gulick, R. (2014). Consciousness. In E.N. Zalta (Ed.), *Stanford encyclopaedia of philosophy.* http://plato.stanford.edu/archives/spr2014/entries/consciousness/
Watt, D.F. (2004). Consciousness, emotional self-regulation and the brain: Review article. *Journal of Consciousness Studies, 11,* 77–82.
Whitehead, A.N. (1985). *Process and reality* [corrected edition]. D.R. Griffin and D.W. Sherburne (Eds.). New York: The Free Press. (originally published 1929)
Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge, Massachusetts: MIT Press.

The Journal of Mind and Behavior
Summer 2014, Volume 35, Number 3
Pages 133–150
ISSN 0271–0137

Development of the Self in Society: French Postwar Thought on Body, Meaning, and Social Behavior

Line Joranger

Telemark University College

The development of the self and behavior toward others were heavily discussed during the French postwar era. According to Foucault, Sartre, and Merleau–Ponty, intersubjective social relations are physical and bodily connections. The physical body is our point of contact with the world, which is a practical world, which we typically engage before any kind of theoretical understanding of what things or people are like. Although there are a number of differences in their ways of thinking concerning the development of the self and social behavior, this paper shows that Foucault and Sartre seem to share Hyppolite's notion that the fulfillment of the absolute self will always be deferred because of an ongoing contradiction in our social behavior.

Keywords: self-consciousness, mental illness, social alienation

The revelation of the underlying ideas that led to discrimination against some groups with regard to both humanity and human rights was one of several reasons that the French postwar intellectual environment included increasing interest in social behavior and the development of the self. De Waelhens (1958) explains the contemporary French interest in social behavior and the development of the self, which he refers to as a body–mind relationship, by the fact that in France, psychoanalysis was paired with phenomenology. According to Spiegelberg (1972), through this pairing, French phenomenology advanced psychoanalysis much more than did psychoanalysis itself. Spiegelberg draws on this context, especially with regard to Merleau–Ponty's body phenomenology, Sartre's existentialism, and Jean Hyppolite's Hegel studies — although, according to Spiegelberg, Sartre, and Hyppolite left their mark to a far lesser extent than

Correspondence concerning this article should be addressed to Line Joranger, Associate Professor, Department of Health and Social Sciences, Telemark University College, Kjølnes ring 56, 3918 Porsgrunn, Norway. Email: line.joranger@hit.no

did Merleau–Ponty. In this context, Spiegelberg seems to have forgotten that in 1954, Foucault published two minor texts about the development of the self, and mental illness, that included a focus on the physical body. These works are the book *Maladie Mentale et Personnalité* [*Mental Illness and Personality*] and Foucault's introduction to the 1954 French edition of the Swiss–German psychiatrist Ludwig Binswanger's 1930 seminal essay on existential analysis, "Traum und Existenz" ["Dream and Existence"], or "Le Rêve et l'Existence," as it was called in the French edition.

In his introduction, Foucault (1954/2001) explicitly makes the connection between psychoanalysis and phenomenology when he claims that phenomenology and psychoanalysis, with regard to Edmund Husserl and Sigmund Freud, contributed to give humankind back its significance and meaningfulness. Although Foucault did not mention Hegel directly, French postwar thinking on the development of the self in society was largely inspired by Hegel's phenomenological and psychoanalytical thinking. As the French epistemologist Georges Canguilhem (1948–1949) wrote, with reference to Hegel, that in a period of world revolution and world war, France discovered a philosophy contemporary with the French Revolution and one that represented, to a great extent, the full realization of the struggle for recognition and the development of the self.

According to Merleau–Ponty, Hegel instituted philosophical modernity. Commenting on one of Hyppolite's Hegel lectures, Merleau–Ponty stated that all the great philosophical ideas of the past century — the philosophies, and psychoanalysis — had their beginnings in Hegel:

> it was he who started the attempt to explore the irrational and integrate it into an expanded reason which remains the task of our century. (. . .) As it turns out, Hegel's successors have placed more emphasis on what they reject of his heritage than on what they owe to him. (1948/1964, p. 64)

Although Hegel, Husserl, and Freud inspired French postwar intellectuals such as Hyppolite, Foucault, Sartre, and Merleau–Ponty, these philosophers did not believe in the historical development of an absolute self, as Hegel did, nor did they support Husserl's notion that one could arrive at an absolute, universal, unhistorical truth through pure phenomenological thinking or Freud's deterministic statement that every psychosis and every mental illness could be traced back to a death instinct or to libido.

To better understand the content of French postwar thought on the development of the self in society and its relation to body, meaning, and social behavior, I will present an outline of Hegel's phenomenological thought as well

as material from the works of Hyppolite, Foucault, Sartre, and Merleau–Ponty.[1] Although there are a number of differences in their ways of thinking about the self, this paper shows that Hyppolite, Foucault, and Sartre share the notion that the fulfillment of the absolute self will always be deferred because of an ongoing contradiction in our social behavior.

Hegel's Phenomenological Thinking on the Development of the Absolute Self

In the *Philosophie des Geistes* [*Philosophy of Mind* (1830/2003)], which is part three in the *Enzyklopädie der Philosophischen Wissenschaften*, Hegel claims that the fight for social recognition is a *life and death struggle* through history that will end in a peaceful political reunion of self-consciousness and reason.

> The fight ends in the first instance as a one-sided negation with inequality. While the one combatant prefers life, retains his single self-consciousness, but surrenders his claim for recognition, the other holds fast to his self-assertion and is recognized by the former as his superior. Thus arises the status of *master* and *slave*. (1830/2003, § 433)

Hegel contends that our social life and the commencement of political union emerge in the battle for recognition under the subjugation of a master. Force, which is the basis of this phenomenon, is not based on rights but rather is a necessary and legitimate factor in the passage from the state of isolated self-consciousness into the state of what Hegel calls the universal self-consciousness. The fulfillment of the absolute self is the affirmative awareness of the self in another self. Each self, as a free individuality, has its own absolute independence but, by virtue of the negation of its immediacy or appetite, does not distinguish itself from the other. Each is thus universal self-consciousness and objective; each has "real" universality in the shape of reciprocity insofar as each knows itself to be recognized in the other free man and is aware of this insofar as each recognizes the other and knows him to be free. The reappearance of self-consciousness is, thus, a form of consciousness that is at the root of all true mental or spiritual life, "in family, fatherland, state, and of all virtues, love, friendship, valour, honour and fame" (§ 436). Hegel believes that the principle of the *free* mind is to make the merely given element (*das Seiende*) in consciousness into some-

[1]There were, of course, several other French postwar writers who were interested in Hegel and the development of the self in society from a phenomenological and psychoanalytical point of view, including Simone de Beauvoir, Jean Wahl, Louis Althusser, Maurice Blanchot, Georges Bataille, Jacques Lacan, Daniel Lagache, and Gilles Deleuze. Later, Félix Guattari, Jacques Derrida, and Julia Kristeva. I have chosen to focus on Hyppolite because he was an important French postwar Hegel interpreter at the time; and on Merleau–Ponty's, Sartre's, and Foucault's early works because of their specific focus on the body–mind relationship.

thing mental (*seelenhaftes*) and, conversely, to make what is mental into a (common) objectivity. Free mind or spirit is to be recognized as the self-knowing truth:

> Free mind stands, like consciousness, as one side over against the object, and is at the same time both sides and therefore, like the soul, a totality. Accordingly, whereas soul was truth only as an immediate unconscious totality, and whereas in consciousness, on the contrary, this totality was divided into the "I" and the object external to it, free mind or spirit, is to be recognized as *self-knowing truth*. (1830/2003, p. 180)

In Hegel's phenomenology, there is the concept of a rational development of a dialectical struggle of social and personal liberation, which will end in the fulfillment of the absolute free human being and the absolute knowledge of truth. Social and political development, according to Hegel, are based on the other's attempts to reduce the other from a subject to a "slave," which, in turn, leads to social anxiety and battles for recognition. Thus, long before Freud, Hegel saw the meaning of dream and imagination. According to Hegel, sleep is a restitutional force, an investigation of our daily activities. To sleep and dream is to return to the general nature of subjectivity (§ 398), which is the substance of psychoanalysis. Nevertheless, it is important to distinguish between the state of dreaming and the state of wakefulness. The person who remains dreaming while awake was, according to Hegel, considered mentally ill. Several years before Freud, Hegel saw mental illness as a relapse into an earlier state of soul development — that is, childhood. This is the unconscious playing with natural instinct, or what he calls emotional life (*Gefühlsleben*) [§§ 403–408].

Like Freud, Hegel did not see mental illness and rationality as opposites but as two interrelated phenomena that share the same underlying structure in which each informs the other in significant ways. The healthy mind grapples with the same sorts of contradictions and feelings of alienation, the same "infinite pain" that characterizes insanity (§ 382). According to Hegel, people with different personalities react differently to their social environment. In this sense, one can observe that one organic being is more sensitive or more irritable or has a greater reproductive capacity than another — just as we observe that the sensibility of one is different from that of another, and people respond differently to a given stimulus (§§ 404–408).

Hyppolite: The Development of the Absolute Self is Forever Deferred

Hyppolite's lectures on Hegel were presented for the students of the École normale supérieure. In attendance at these lectures were Sartre, Merleau–Ponty, and Foucault. His lectures called forth questions about psychoanalysis and the logic of passion, mathematics, the formalization of discourse, and information theory and its application. The lectures explored questions about an existence

that constantly associates and dissolves its relationships. They explained the interaction between the self and the other as a perpetual existential conflict with no democratic progress. In this vein, Hyppolite read Hegel in sharp contrast to the more Marxist-oriented interpretations of another famous Russian–French Hegel interpreter of the time, Alexandre Kojève. Unlike Kojève, Hyppolite described the subject of Hegel's theories as a tragic component of human existence. Although both Hyppolite and Kojève argued for the historical dimension of the subject's temporality, Hyppolite's history of philosophy had no human components and no notions of the subject as an historical actor.

Despite his anti-existentialist view of the human subject and although his theory did not have a specific focus on the human body, Hyppolite read the relationship to the human experience in the manner of most other French existentialists: as a struggle for recognition. Like Hegel, he saw this struggle in relation to the desire to be held in high esteem. However, in *Genèse et Structure de la Phénoménologie de l'Esprit de Hegel* [*Genesis and Structure of Hegel's Phenomenology of Spirit* (1946/1974)], Hyppolite suggested that there is no question of an historical dialectic of recognition evolving as Hegel described it in the *Philosophy of Mind*. Seen in this way, the completion of the Absolute seems forever deferred (Hyppolite 1946/1974, p. 145). There is no rational, dialectical struggle of liberation in which the oppressed enlightens the overlord and vice versa. Life leans more in the direction of what Kierkegaard describes as self-agitation, anxiety, and suffering. For Hyppolite, the alienation of subjectivity means that one never agrees with oneself because one continually becomes another in the endeavor to be oneself:

> The self never coincides with itself, for it is always other in order to be itself. It always poses itself in a determination and, because this determination is, as such, already its first negation, it always negates itself to be itself. It is human being "that never is what it is and always is what it is not." (p. 150)

Thus, the finite subject is not limited in the way that an object can be limited. An object does not know its own limit, which is external to it. The subject continually seeks to transgress its limit; it tends toward the infinite, the unconditioned. This understanding (*Verstand*) is reason (*Vernuft*), but by the same token, it transgresses the very sphere of objects. This infinite is not an object; it is a task whose accomplishment is forever deferred. According to Hyppolite, it is no longer the concept of reason that regulates experience but that of the idea and the infinite practical task in relation to which all knowledge and all knowing are organized. Because the subject always fails in its endeavor to become whole and united, its basis remains always, Hyppolite suggests, in an unhappy consciousness (p. 191). The experience of the self becomes inadequate and incomplete and ceases to correspond with the objects of truth, and our knowledge of death enforces our knowledge of limited time. In the encounter with others,

we learn that the self does not exist all at once but is alternately lost and then recovered. Concretely, this is the very essence of human beings. They are never what they are; they always exceed themselves and are always beyond themselves; they have a future; and they reject all permanence except the permanence of their desire, which is aware of human beings as desire.

According to Judith Butler (1999), it might seem that Hyppolite's vision of death has engaged Freud's (1922) vision in *Beyond the Pleasure Principle* and that all desire is, in some sense, inspired by a fundamental striving toward death (i.e., the desire to die). However, Butler believes that both Hegel's and Hyppolite's Christianity imply that death, to which consciousness aspires, is itself a fuller notion of life.

> Following Kojève and Jean Wahl, Hyppolite restricts himself to the interpretation of death offered in the section of lordship and bondage. He takes seriously the facticity of the body, finitude as the condition of a limited perspective, corporeality as a guarantor of death. The vision of a new life, a life beyond death, remains purely conjectural in Hyppolite's view, but it is a conjectural that holds sway in human life. (1999, p. 91)

Thus, the fact that we never will conform to another human being and the fact that intersubjective forms of cohabitation cause considerable interaction challenges for the development of the self are the lived experience of the infinite. "To cultivate oneself is not to develop harmony, as in organic growth, but to oppose oneself and rediscover oneself through a rending and a separation" (Hyppolite, 1946/1974, p. 385).

Sartre: The Bodily Self for Others — the Bodily Self for Itself

Like Hyppolite and Hegel, Sartre concerns himself with the development of the self in society. In *L'Être et le Néant* [*Being and Nothingness* (1943/2003)], Sartre contends that we first and foremost meet others as rival consciousnesses, as rival sources of freedom and power. He suggests that our relationships with others are *intersubjective* in the sense that we, in our development of the self, are dependent on others' judgment. This is something we fear and would prefer to escape. In this context, we are, like Hegel, talking about an intersubjective interaction that leads to a predictable interaction of dominance and submission in which we either attempt to overpower the other (the sadistic strategy) or to surrender to the command of the other's mastery (the masochistic strategy). In both cases, Sartre believes, we confirm that there is a need for us, that we are powerful, and that we are substantive. If these strategies are not successful, we have a third option: to withdraw from all relationships to avoid the threat of the other's *gaze*, which can destroy us.

Sartre exemplifies the gaze by relating the experience of a jealous person who observes the other through a keyhole (1943/2003, p. 282ff). The observer enjoys

the feeling of having the body of the other, who does not know that she is observed as an object in his power. However, what the observer does not know is that he is also an object of observation by a third person while he is spying in the keyhole. This sensation of suddenly being discovered promotes, according to Sartre, a feeling of shame that is perceived as humiliating because the observer himself is reduced from being the one who observes and has power to being the observed (the object of the other) at the mercy of a negative judgment from others. Nevertheless, the sensation of being discovered by another leads us from the unreflective consciousness for itself in isolation to the reflective consciousness in the world of others. "It is shame or pride which reveal to me the Other's look and myself at the end of that look. It is the shame of pride which makes me *live*" (pp. 284–285).

Because human relationships are always based on one's attempts to reduce the other as a subject to an object, Sartre believes an equal "we" is impossible to achieve. Thus, according to Sartre, it is understandable that one wants to withdraw from social communication. However, like Hegel and Hyppolite, he argues that this is an impossible solution in the long run. We are nothing if we are not in an intersubjective relationship with the other. In this sense, the other is a necessary source of affirmation of one's own existence. Wanting to rise above this connection is the same as signing one's own death warrant.

In contrast to contemporary existence-phenomenalism, in *L'Existentialisme est un Humanisme* (1946) [*Existentialism Is a Humanism*], Sartre claims that man, in this intersubjective situation, is condemned to be *free* because he has *not* created himself but is still free: "From the moment that he is thrown into this world he is responsible for everything he does" ("*parce qu'une fois jeté dans le monde, il est responsible de tout ce qu'il fait*") [p. 40]. He suggests that man is not only what he conceives himself to be but also what he wants to be. We are nothing but that which we make of ourselves (pp. 29–30). No a priori morals, values, or injunctions exist to support us in life, as Kant and Husserl claim, and there is no materialistic or libidinous determinism or rational social development, as Marx, Freud, and Hegel claim; instead, man is freedom.

In *Being and Nothingness*, Sartre suggests that our basic anxiety is related to the awareness of freedom. Each choice is associated with anxiety because every choice commits us and makes us responsible not only for ourselves but also for others. In this respect, we live constantly in relation to a desired *future* through our projects, expectations, beliefs, and desires (Sartre, 1943/2003, pp. 147–152). How we perceive and relate to our society and our social situation here and now is determined by our desires for the future, such as our wish that our dearest friend will come home from Berlin at any minute. Sartre entirely overturns the assumption that my *choices* are determined by who I am and make me who I am. What we are and what we become are entirely dependent on our choices and intentions. Our responsibility for who we are is therefore total; we can set

ourselves free with regard to our future, and we will have to make the choices that will determine our future.

In *L'imaginaire Psychologie Phénoménologique de l'Imagination* [*The Imaginary: A Phenomenological Psychology of the Imagination* (1940/2004)], Sartre emphasizes the tendency to create internal images of ourselves and who we want to be or become. He believes, like Hegel, that it is important to distinguish between objects in the real world (the physical body as such) and objects in the fictional quasi-world (the body as an imagination). Perceptions and imaginations offer not only an escape from a specific and undesirable situation but also an "(. . .) escape from all the constraints of the world [;] they seem to be presented as a *negation* of the condition of being in the world, as an anti-world" (1940/2004, p. 136). By claiming this, Sartre confronts psychoanalysis by rejecting the notion that something that involves consciousness can also be unconscious. He does not deny that the unconscious exists — only the notion that the unconscious is a place where mystical and meaningful things happen outside of consciousness.

For Sartre, the body is the sediment of the past that we project toward the future. It is that whose surface power is inscribed and that by whose powers such power is "incorporated." It is the natural symbol as well as the existential basis of culture. In part three of *Being and Nothingness*, Sartre dedicates a full chapter to what he believes to be a three-dimensional body as such — that is, the body in relation to society or *for others*, the body in relation *to itself*, and, lastly, the body in relation to an *ontological* notion. Because the body in Heidegger's terms is "being-in-the-world" and because the body is our being-there in the world, any description of the body has as its correlate a disruption of the world — that is, in Husserl's terms, the *Lebenswelt*, the life-world.

Sartre claims that the body as being-for-others is a body in a social situation. In this case, the other's body is meaningful and is not perceived as a thing among things, as if it were an isolated object with purely external relations with other objects (*objets*). He suggests that in this context, there is a radical difference between objects and human beings. Suppose that we see a man in a public park:

> If I were to think of him as being only a puppet, I should apply to him the categories which I ordinarily use to group temporal–spatial "things." (. . .) Perceiving him as a *man*, on the other hand, is not to apprehend an additive relation between the chair and him; it is to register an organization *without distance* of the things in my universe around that privileged object. (1943/2003, p. 278, italics in the original)

Although, in this example, the other is a body by virtue of the fact that I am looking at him and not vice versa, the other is, according to Sartre, perceived as a situated object around whom society is organized. The other's body is seen as a center of his own fields of perceptions and actions, and the space he inhabits

is the space in which he lives. As already noted, this interpretation indicates two dimensions of the body: the body as being-for-itself (my own body as it is normally for me) and the body as being-for-others (my body as it normally appears to others or, equivalently, the body of the other as it normally appears to me). A third ontological dimension is then generated, so to speak, by the interaction between these first two dimensions: "My awareness of being an object for others means that I also exist for myself as body known by the others" (p. 375).

Frie (1997, p. 60) suggests that although Sartre conflates the ontological dimension of *Mitsein* with the experience of a we-subject, Sartre makes an important point: being-with-others follows from being-for-others. In this, Sartre identifies a dimension of affectivity, revealed, for example, by illness, which he calls "my body on a new plane of existence" (a psychic body) (1943/2003, p. 361), and an "aberrant type of appearance" when my own body appears to me as one object among other objects (p. 377–381). According to Sartre, this objectifying of the body happens when, for example, the doctor looks at my body as a physical object. At that point, "my body is designated as alienated" ("*Mon corps, en tant qu'aliéné*") (p. 376; 1943, p. 393). The experience of social alienation is achieved in and through affective structures, such as shyness, blushing, and sweating. He describes these feelings as a constant consciousness not of the body as being-for-itself but of the body as being-for-others. Sartre believes this constant uneasiness, which is the apprehension of my body's social alienation as irremediable, can determine psychoses such as ereuthophobia (a pathological fear of blushing), which are merely the horrified metaphysical apprehensions of the existence of my body for others (1943/2003, p. 376). He suggests that the explanation here is that we attribute to the body-for-other as much reality as we do to the body-for-us — or, more accurately, the body-for-other *is* the body-for-us, but it is inapprehensible and alienated. It appears to us that the other accomplishes for us a function of which we are incapable but that nevertheless is incumbent on us: to see ourselves as we are.

Merleau–Ponty: The Holistic Structure of Body, Meaning, and Behavior

In Merleau–Ponty's concept of the development of the self in society, man is not an object of his surroundings. Merleau–Ponty's relation to the environment is not objective in the sense of something unambiguous and measurable; the human body is not a type of machine, as Descartes suggests. By claiming this relation, Merleau–Ponty rejects Sartre's metaphysical and Cartesian dualism of the body — that is, the body as being-for-itself (my own body as it is normally for me) and the body as being-for-others (my body as it normally appears to the other or, equivalently, the body of the other as it normally appears to me), which is an observable, physical body in society with others. According to

Merleau–Ponty, if I thought of myself and my body in this vein, I would not call it mine and I would not be me. He suggests that it is better to say "I am my body" — that is, my meanings are found in the structures of my body's behavior, and it is the center of the world in which I exist.

In *Phénoménologie de la Perception* [*Phenomenology of Perception* (1945/2002)], Merleau–Ponty suggests that one cannot speak of different realities and different self-consciousnesses or body awarenesses. He refers to an example in which an organist, despite the fact that he plays on a new and unknown organ, soon becomes so familiar with the organ's characteristics that it cannot be explained by mechanical learning and adaptation. In contrast to Sartre, who rejects that unconsciousness represents meaningful things outside of consciousness, Merleau–Ponty demonstrates that it is because of the sub-consciousness of the bodily self that the organist "installs" (*installe*) himself, so to speak, in the organ and creates an existential self in relation to the musical instrument (1945/2002, pp. 167–168). Our intersubjective social relations with others are thus a physical and bodily connection, which is crucial for understanding ourselves in relation to society and to other people.

> True reflection presents me to myself not as idle and inaccessible subjectivity, but as identical with my presence in the world with others, as I am now realizing it: I am all that I can see, I am an *intersubjective* field, not despite my body and historical situation, but, on the contrary, by being this body and this situation, and through them, all the rest. (p. 525)

Merleau–Ponty suggests that the essential characteristic of the self is that the body is, or has, a pre-objective relationship with its surroundings. This relationship has intentionality in Kant's and Husserl's sense of the word in that the body is directed toward comprehending society. Herein resides the title and significance of his work *Phenomenology of Perception*. The "phenomenon" is what comes into view; like Husserl, Merleau–Ponty wants to regard the phenomenon carefully and without prejudice. What stands out for a trained phenomenologist is a perceptual field that opens up the perceptual body, and this area contains many layers of meaning. In the first layer are the pre-objective phenomena themselves. These phenomena are open, ambiguous phenomena to which the human body responds. The body and its surroundings constitute an internal relational structure in which the two elements mutually refer to each other. This structure is the meaning of Heidegger's concept of being-in-the-world, which Merleau–Ponty refers to as being-to-the-world (*être au monde*). By showing how the human body is not mechanically, biologically, or intellectually related to the world but, rather, is existentially related to it, Merleau–Ponty outlines a new way of examining and reinterpreting the body–mind

relationship that extends beyond Heidegger's notions, which do not fully examine the body–mind relationship and the problem of perception.[2]

Although we find many of Merleau–Ponty's arguments about the body–mind relationship again in Sartre's perceptions of the self, his theory does not relate the body specifically to perception, even if he (like Husserl) believes that the body is present in every perception. When Sartre speaks of the position and movement of the body, he refers neither to a spatial object's motion nor to a position in a geometric room. The spatiality of the body is not linked to a position but to a situation. The body is not a point among others; rather, it is the anchor in the world that makes all other coordinates possible. In other words, the body's "here" is an absolute "here" as opposed to the place where I currently find myself. There can never be a "there" for me.

According to Merleau–Ponty, it is important that we move beyond the natural world and rediscover the social world, not as an object or sum of objects but as a permanent field or dimension of existence. Our relationship to the social, like our relationship to the world, is deeper than any express perception or judgment. It is as false to place ourselves in society as an object among other objects as it is to place society within ourselves as an object of thought. In both cases, the mistake lies in treating the social as an object. In contrast to Sartre, Merleau–Ponty argues that our identity and our behavior are presented in such a fundamental and profound way that we only explicitly become aware of them when our usual interaction with society is disturbed by something that is forced upon us, such as mental illness, and in situations similar to what Jaspers calls *boundary situations* (*Grenzsituationen*), which constantly affect our psychic and physical lives (Jaspers, 1932/1971, vol. II, chapter 7). If we attempt to escape boundary situations by managing them with rationality and objective knowledge, we must necessarily flounder. Instead, boundary situations require a radical change in attitude in one's normal ways of thinking. The proper way to react within boundary situations is, according to Jaspers

> not by planning, and calculating to overcome them but by the very different activity of *becoming the Existenz we potentially are*; we become ourselves by entering with open eyes into the boundary situations. We can know them only externally, and their reality can only be felt by *Existenz*. To experience boundary situations is the same as *Existenz*. (1932/1971, p. 179, italics in the original)

---

[2]By placing the body consciousness before the mind consciousness, Merleau–Ponty approaches radical behaviorism, which asserts that the human psyche cannot be examined and, thus, that only external and visible behavior remains as the subject of science. Merleau–Ponty himself was aware of the similarity between his work and behaviorism. In *La Structure du Comportement* [*The Structure of Behavior* (1942/2011)], Merleau–Ponty claims that behaviorism and Pavlov's reflexology misinterpreted existence by understanding it in response to stimuli, analogous with the mechanistic cause–effect relationships between objects; see chapter II. "Higher form of Behavior" (pp. 52–128).

In *La Structure du Comportement* [*The Structure of Behavior* (1942/2011)], Merleau–Ponty suggests that we can find in the disintegrated consciousness an illustration of the mind–behavior parallelism in which conscious states run parallel to isolated bodily occurrences. In such a sickness, this isolated body may causally affect our perception so that what I perceive may serve as a subjective veil between me and the real things around me in society. However, the mind and the body of the integrated person are not allowed to disintegrate in this way. This person's body does not act as a separate cause to introduce distortions into his perceptions. A disintegrated self-consciousness may be parallel to an isolated cycle of physical events, but true consciousness is parallel to society and can hardly be explained logically or by scientific concepts (p. 224).

Like Sartre, Merleau–Ponty's anti-deterministic view opposes Freud's attempts to diminish the human psyche into mere sexual desire. He does not believe that anything in the human psyche can be reduced to standardized or logical categories. We always exist within the world, or in *situations*. Referring to Freud's psychoanalysis, Merleau–Ponty suggests, "there is no explanation of sexuality which reduces it to anything other than itself, for it is already something other than itself, and indeed, if we like, our whole being." (1945/2002, p. 198)

Foucault's 1954 View on Body, Meaning, and Social Behavior

In significant contrast to his later works and to contemporary existentialist thought, in *Maladie Mentale et Personnalité* (1954),[3] Foucault approaches the self by distinguishing between what one can scientifically explain — that is, our physically observable body — and what one cannot scientifically explain — our minds and our inner psychological feelings and perceptions. He presents two approaches to the social self and the mind–body–behavior relationship: a phenomenological, interpretative, non-scientific approach and an explanatory, scientific, neurological approach.

In the first part of *Maladie Mentale et Personnalité*, using the phenomenological first-person perspective, Foucault refers to phenomenological psychiatrists such as Binswanger, Kuhn, Séchehaye, and Minkowski to offer examples of a self-consciousness that would seem unrecognizable for most people but that has become real for the mentally ill person. This self-consciousness is, according to Foucault, effected by the fact that the body often ceases to be a point of reference against the opportunities in the world ("*Le corps cesse alors d'être ce centre de reference autour duquel les chemins du monde ouvrent leurs possibilités*") [1954, p. 65]. The body becomes unrecognizable to consciousness because its impulses stem from a mysterious exteriority. Foucault refers to one of Minkowski's patients, who describes how he experiences his body as a body hard as wood, as a body hard

[3]This edition is not translated into English. All translations are mine.

as brick, a body black as water, where the teeth are perceived as ends in a drawer made of hard oak tree (p. 66) Occasionally, according to Foucault, we see that full body awareness (that is, the awareness of a physical body in time and space) disappears to the extent that one ultimately has only an awareness of a disembodied life and an unrealistic idea of an immortal existence.

In the fifth and final chapter of *Maladie Mentale et Personnalité, "La Psychologie du Conflit"* ("Conflict Psychology"), Foucault uses examples from Pavlov (pp. 91–102) to turn away from his earlier phenomenological view and to offer an explanatory, neurologic, socio-cultural approach to the self and the mind–behavior relationship. Foucault suggests that the consequence of mental illness and alienation is that the bodily nervous system, in "natural" and bioneurological ways, transforms sociocultural conflicts and historical development (the present historical condition) into inner personal life histories, which he believes can lead to paradoxical defense reactions. Foucault emphasizes that Pavlov's most important contribution to psychology was his study of how external stimuli and environmental conditions can trigger internal anxiety reactions and schizophrenic experiences of self.

Because Pavlov's research showed that the nervous system as a whole normally manages to balance environmental impacts, Foucault's concern is directed toward how these seemingly normal nerve functions can be the cause of pathological activities (1954, p. 94). If a person's central nervous system is subjected to a strong activation (*excitation*), such as violent agitation, Foucault suggests that this will inductively be followed by an inhibitory defense reaction (*inhibition*), followed by a blocking and a corresponding strengthening of the nerve cells' excitation and inhibition. In normal cases, this reaction will inductively lead to the reduction and eventual cessation of the process. For the mentally ill person, however, the process will only continue in an ongoing cycle. Foucault suggests that in these cases, one can release emotional stress through an organic lobotomy, but he believes that this does not change the patient's interior work (p. 108). As psychoanalysts do, one can also address a current conflict by appealing to subtle instincts and past events. However, according to Foucault,

> When we know that the disease always refers to a dialectical conflict situation, there will be both efficient and functional treatment that takes place in this particular situation. (*Et d'un autre côté, puisque la maladie se réfère toujours á une dialectique conflictuelle d'une situation, la thérapeutique ne peut prendre son sens et son efficacité que dans cette situation.*) [pp. 108–109]

He asks,

> If the subjectivity of insanity is both a call to and an abandonment of the society, is it not the society itself we should ask the secret of its enigmatic status? *(Si cette subjectivité de l'insensé est, en même temps, vocation et abandon au monde, n'est-ce pas au mone lui-même qui'il faut demander le secret de cette subjectivité énigmatique?)* [p. 69]

By asking this question and by turning his attention to the external social environment, Foucault turns away from Husserlian and phenomenological inward analysis. With reference to the mentally ill person's difficulty with social affiliation and dialogue, he suggests that a whole social evolution was required before dialogue could become a mode of human interaction (pp. 27–28). This evolution was made possible only by a transition from a society immobile in its hierarchy of moment, which authorized only order, to a society in which the equality of relations enabled and ensured potential exchange, fidelity to the past, engagement in the future, and reciprocity of points of view. Foucault asserts that the patient who is incapable of dialogue regresses through this social evolution; dialogue, as the supreme form of the evolution of language, is replaced by a sort of monologue (p. 28). By losing the ambiguous potentiality of dialogue, the patient loses mastery over his symbolic world and the ensemble of words, signs, and rituals — in short, all that is allusive and referential in the human world. Seeing Foucault's concept of evolution in light of Hegel's master–slave dialectic, the mentally ill person or the person with social fear will, thanks to modern social and democratic developments, both confirm and increase his status as a slave in relation to other more adaptively social individuals.

Like most of his contemporaries, Foucault criticizes Freud's psychoanalysis for camouflaging the unique expression of illness and what he believes to be the authentic and existential dream language, or "dream meaning." He wants to free the analysis of pathological regression from the myth that mental illness is related to a certain psychological meaning (such as Freud's libido), which is seen as the raw material of evolution and which, progressing in the course of individual and social development, is subject to relapses and can revert, through illness, to an earlier state. According to Foucault, we must accept the specificity of the morbid (archaic-like) personality as strictly original. Therefore, the analysis must be conducted further, and this evolving, potential, and structural dimension of mental illness must be completed by the analysis of the dimension that makes it necessary, meaningful, and historical (p. 35).

In his introduction to the 1954 French edition of Binswanger's essay "Traum und Existenz," Foucault (1954/2001) suggests that the essential function of dream analysis is less to revive the past than to make declarations about the future. Such an analysis anticipates and announces the moment at which the patient will finally reveal the secret that she does not yet know and that is nonetheless the heaviest burden of her present. The dream anticipates the moment of liberation to come. It is a prefiguring of history even more than it is an obligatory repetition of the traumatic past. According to Foucault, man has known since antiquity that in dreams, man encounters what he is and what he will be, what he has done and what he is going to *do*, discovering there the knot that ties his freedom to the necessity of the world (1954/2001, p. 113). He criticizes Sartre, who, like Hegel, distinguishes imagination from reality.

According to Foucault, to become totally free from social battles and social restrictions, we must recognize that our imaginative life is as real as our "lived" life (pp. 138–139).

French Postwar Thinking of the Development of the Self in Society

With regard to French postwar thought on the development of the self in society and its relation to body, meaning, and social behavior, like Hyppolite, Foucault thinks that modern intersubjective forms of cohabitation have caused considerable interaction challenges for the development of the self in general and for social behavior specifically. With regard to mental illness and antisocial states, dialectical recognition may not be the central issue because life becomes synonymous with anxiety, fear, and distress. In the same manner as Hyppolite, Foucault's notion in *Maladie Mentale et Personnalité* seems to be that the social alienation of the self means that one can never match oneself. The development of the self in society becomes, in contrast to Hegel's view, inadequate, incomplete, and out of sync with the objects of truth.

Because the subject always fails in its endeavor to become whole and united, the basis of self remains, as Hyppolite (1946/1974, p. 191) describes it, always an unhappy consciousness. The development of history is, from this perspective, not rational and liberating, as Hegel and (in many ways) Marx would argue; rather, it is irrational and oppressive because it locks man into specific positions of interaction and patterns of behavior that prevent him from playing out his personal and existential expressions and imaginations. Foucault seems to believe that this type of social anxiety occupies the behavior of the person with mental illness to such an extent that he stops communicating with other people. By withdrawing from all social intercourse, a person with mental illness escapes not only from himself but also from the other's gaze, with the result that he makes his situation even worse by ensuring that the people around him perceive him as an alien (a slave) in his own universe.

In the same vein, although he does not share Foucault's dual methodological approach to the self (that is, a subjective phenomenological focus on the mind and a reflexological and socio-cultural focus on the body), Sartre explains the alienated self by dividing self-consciousness into three parts, the body as being-for-itself, the body as being-for-others, and the body as an ontological dimension: "My awareness of being an object for others means that I also exist for myself as body known by the others" (1943/2003, p. 375). According to Sartre, the self can only be released from the burden of being the object of the other's gaze and judgment by imagining a freedom and an identity in which everything says "I," fully incorporated in a quasi-world. Opposed to this view, Merleau–Ponty and Foucault (in his introduction to "Traum und Existenz") seem to believe this "quasi-world" is connected to reality itself. The self — that

is, the mind, behavior, and surroundings — constitute, in this case, an internal relation structure in which they mutually refer to each other.

Like Foucault, Merleau–Ponty's notion is that our imaginations, movements, dreams, and language represent the situation itself — nothing unreal. For Merleau–Ponty, as for Foucault, one cannot create a scientific method that can study the life of signs within society. Rather, Merleau–Ponty suggests that our existential signs cannot be reduced to a set of facts that are capable of being reduced to others or to which they can reduce themselves. There can be no objective science of subjectivity. We are all that we are on the basis of a de facto situation that we appropriate for ourselves and that is ceaselessly transformed by a sort of escape that can never be an unconditioned freedom.

Merleau–Ponty's holistic notion is that one cannot understand the development of the self in society by using different views and methods to understand and explain the mind and body separately, as Foucault did in *Maladie Mentale et Personnalité*. For Merleau–Ponty the self will not be experienced as a self if one divides it into separate parts because I am my mind and my behavior; my mind and behavior are the center of the world in which the self exists and cooperates with other selves. From this perspective, one cannot speak of a socially alienated self or assert that the development of an absolute self is forever deferred because the self is always absolute in relation to itself and others, even with regard to illness.

Unlike Hyppolite's concept of the self as an unhistorical actor, Foucault, Sartre, and Merleau–Ponty demonstrate that the self is always in a specific historical and cultural setting searching for subjective and collective meanings. According to these authors, intersubjective social relations involve a historical, physical, and bodily connection that is crucial for understanding ourselves in relation to others. For Foucault and Sartre, the social self seems to be both psychologically and physically active and influential, alienated, and restricted. Because the natural and social self will always resist becoming an object of its surroundings, the self will always strive for development and integration with the world of the other. In this vein, because we cannot escape the judgment of others if we want to become real (cf. Sartre), we shape an illusion of invulnerability. Merleau–Ponty and Foucault claim that this imagination is part of being human and life itself.

References

Binswanger, L. (1930). *Traum und Existenz*. Zürich: Girsberger.

Butler, J. (1999). *Subjects of desire: Hegelian reflections in twentieth-century France*. New York: Columbia University Press.

Canguilhem, G. (1948–1949). Hegel en France. *Revue d'Hisloire el de Philosophie Religieuses, 28–29,* 282–297.

De Waelhens, A. (1958). *Existence et signification*. Louvain Paris: Nauwelaerts.

Foucault, M. (1954). *Maladie mentale et personnalité*. Paris: Presses Universitaires de France.

Foucault, M. (2001). Introduction. L. Binswanger, *Le Rêve et l'Existence*. In D. Defert, F. Ewald, and J. Lagrange (Eds.), *Dits et écrits I, 1954–1976* (pp. 93–147). Paris: Gallimard. (originally published 1954)

Freud, S. (1922). *Beyond the pleasure principle* [C.J.M. Hubback, Trans., second edition]. Vienna: International Psycho-Analytical Press.

Frie, R. (1997). *Subjectivity and intersubjectivity in modern philosophy and psychoanalysis: A study of Sartre, Binswanger, Lacan, and Habermas*. London: Rowman & Littlefield.

Hegel, G.W.F. (2003). *Philosophy of mind* [W. Wallace and A. V. Miller, Trans.]. In *Encyclopaedia of the philosophical sciences, part three* [reprinted edition.]. Oxford: Clarendon Press. (originally published 1830)

Hyppolite, J.G. (1946). *Genèse et structure de la "Phénoménologie de l'Esprit" de Hegel*. Paris: Éditions Montaigne.

Hyppolite, J.G. (1974). *Genesis and structure of Hegel's* Phenomenology of Spirit [S. Cherniak and J. Heckman, Trans.; J.M. Edie, Ed.]. Evanston: Northwestern University Press. (originally published 1946)

Jaspers, K. (1971). *Philosophy* [E.B. Ashton, Trans.]. Chicago: University of Chicago Press. (originally published 1932)

Merleau–Ponty, M. (1964). *Sense and non-sense* [H.L. Dreyfus and P.A. Dreyfus, Trans.; J. Wild, Ed.]. Evanston: Northwestern University Press. (originally published 1948)

Merleau–Ponty, M. (2002). *Phenomenology of perception* [C. Smith, Trans.]. London: Routledge. (originally published 1945)

Merleau–Ponty, M. (2011). *The structure of behavior* [A.L. Fisher, Trans.]. Pittsburgh: Duquesne University Press. (originally published 1942)

Sartre, J.-P. (1943). *L'être et le néant: Essai d'ontologie phénomenologique*. Paris: Gallimard.

Sartre, J.-P. (1946). *L'existentialisme est un humanisme*. Paris: Gallimard.

Sartre, J.-P. (2003). *Being and nothingness: An essay on phenomenological ontology* [H.E. Barnes, Trans.]. London: Routledge. (originally published 1943)

Sartre, J.-P. (2004). *The imaginary: A phenomenological psychology of the imagination* ["Routledge," Trans.]. London: Routledge. (originally published 1940)

Spiegelberg, H. (1972). *Phenomenology in psychology and psychiatry: A historical introduction*. Evanston: Northwestern University Press.

The Journal of Mind and Behavior
Summer 2014, Volume 35, Number 3
Pages 151–166
ISSN 0271–0137

Expressivism, Self-Knowledge, and Describing One's Experiences

Tero Vaaja

University of Jyväskylä

In this article, I defend an account of self-knowledge that allows us a considerable first-person authority regarding our subjective experiences without invoking privileged access. I examine expressivism about avowals by contrasting it with "detectivist" and "constitutivist" accounts of self-knowledge, following the use of these terms by David Finkelstein. I proceed to present a version of expressivism that preserves some of the valid motivating insights of detectivism and constitutivism as essential parts. Finally, I point out how my account views self-knowledge as a cognitive and conceptual ability that can be cultivated; the account construes self-knowledge as a process.

Keywords: expressivism, first-person authority, avowal

Each of us is normally the best person to ask when it comes to our own feelings and experiences. Speaking about one's own mental states is generally held to carry a special epistemic authority. Moreover, this authority belongs exclusively to the first person; others are not admitted to have a similar claim to know someone's experiences even if they are extremely well-informed and familiar with them. I take these to be facts on first-person authority as they appear in the practice of human life quite universally.

Such authority has a central place in social life; denying it can easily (and legitimately?) be taken as an offence. However, it might be that philosophers have historically been overconfident about the special security of our knowledge of our own minds. Carruthers (2011) argues that self-knowledge is interpretive and prone to confabulation. Schwitzgebel (2011; Hurlburt and Schwitzgebel, 2007) claims that we might be regularly wrong about even quite fundamental features of our conscious experience. Therefore it is important to be clear

Correspondence considering this article should be addressed to Tero Vaaja, Yhteiskuntatieteiden ja filosofian laitos, PL35 (Ylistönmäentie 33), 40014 Jyväskylän yliopisto, Finland. Email: tero.t.vaaja@jyu.fi

about the nature of first-person authority, and the conditions in which it may be legitimately challenged.

In this article, I seek to give a modest account of self-knowledge that still respects the special status of the subject as a knower of her own mental states. I treat commonsensical first-person authority as an explanandum, setting aside accounts that seek to dethrone the notion altogether. I start by presenting two contrasting views about the nature of self-knowledge and the basis of first-person authority. I point out how each of these views, "detectivism" and "constitutivism," is unsatisfactory and how expressivism about avowals, an idea inherited from Wittgenstein (1953), can be seen as preferable to them. I owe the terms detectivism and constitutivism, as well as the main drift of the argument in the first half of this paper, to Finkelstein (2003). Another way to refer to these two contrasting views would be to call them (species of) empiricism and rationalism about self-knowledge, as is done in Gertler (2011). I proceed to present a version of expressivism that incorporates some of the good insights made by detectivism and constitutivism. As explained in the conclusion, I hope my view to be meritorious in respecting commonsensical first-person authority without invoking privileged access, i.e., an idea of a special epistemic channel that makes self-knowledge unproblematic to come by. I also seek to do justice to the meaning of "self-knowledge" as a process that has to do with the personal development of one's conception of oneself.

Detectivism

What is it that makes psychological self-ascriptions, or avowals, especially secure?[1] One way of answering is to appeal to introspection, combined with some form of privileged access. The idea is simple: people come to know what their own mental states are like because they are the ones who directly feel or perceive those states. We are assumed to have an "inner sense," or some naturally evolved capacity that enables us to inwardly monitor our mental states. These are forms of what Finkelstein (2003) calls detectivism: the view that the source of self-knowledge is a perceptual or quasi-perceptual act of detecting that allows us to find out our own mental states.

So, one possible explanation for first-person authority is a combination of two ideas: first, there is a special way of detecting one's own mental states; and second, that way of detecting is remarkably reliable. Maybe subjects are not completely infallible about everything that goes on in their conscious experience,

[1]I will use "avowal" as an umbrella term to refer to any sincere utterance whereby the subject speaks about her mental condition. This liberal use is not a standard one. According to more restricted uses of the term, what I will later refer to as primitive avowals and intellectual self-ascriptions would not necessarily qualify as avowals.

but they have such a propensity of being right about those things that it cannot be paralleled by any other person.

It is hard to deny that in an obvious sense, the subject of a painful sensation is in a better position to observe that particular pain than anyone else. But it is still far from obvious that this is what grounds the typical way in which first-person authority is granted to subjects, or if this is a good account of what self-knowledge is. Next, I attempt to illustrate the issue by an example; my chosen example in this paper will be a case of describing a sensation of pain.

Example 1

> I have an abdominal pain that I need to describe to a physician. I am able to point out its location and give an evaluation of its intensity on a scale of 1 to 10. I will also describe its qualitative character by a few adjectives. After careful consideration and some effort to find the right words, I say (at time t_1) that my pain is located about ten centimeters up from my waistline, on the left side of my middle abdomen, its intensity is 6, and it is stinging, sharp, distressing, and penetrating.
>
> When I have finished giving my description, I overhear the word "rip," or someone suggests it to me. I say (at time t_2): "Ripping! Yes. That's what my pain is like. That's right; I could not come up with it myself."

When I eventually say that my pain is ripping, I presumably say it with first-person authority. The fact that I needed help in finding the word might give reason for an interlocutor to not take it completely at face value; a question like "Are you sure that is the right word?" might be justifiable. But if I say sincerely and after careful consideration that "ripping" describes my pain perfectly, it is unclear what could ground the claim of someone who insists that I must nevertheless be wrong. In this kind of a situation, any doubt that another person might harbor about the appropriateness of my pain-description will more naturally target my adeptness in the use of the word, rather than the accuracy of my introspective act.

According to detectivism, my statements about my pain are based on perceptual or quasi-perceptual observing. In this case, I am supposedly monitoring my sensation of pain and detecting a ripping quality in it. But detectivism makes it hard to see why my eventual description of my pain as ripping should carry any special authority. It was, after all, based on the same introspective observation that I had already done at t_1, without at that time judging my pain to be ripping. We can make the example clearer by stressing that my sensation of pain stays the same from t_1 to t_2: I am not judging my pain to be ripping at t_2 because it started as non-ripping and then suddenly turned into ripping. Someone could suggest that at t_1 I did not attend to the pain as completely as I did at t_2; the suggestion could be that upon hearing the word "rip," I introspectively probed the pain again to see if the new word fits it, and found a novel ripping quality in it. But it is possible that I would sincerely deny that too, and

testify that my pain features in my experience exactly in the way as it did at t_1. The quality that made me describe it as ripping was in my awareness from the start; I merely came up with a better description of it.

I think it is fairly plausible that in this situation, where I explicitly admit that I do not derive my eventual pain-description from any distinct introspective act, few people would feel that the authority of my avowal diminishes from t_1 to t_2. This suggests that detectivism is not adequate to explain the basis of first-person authority.

Maybe we should waive the detectivist idea and state that inward perceptions are not the source of the authority of my avowals. Instead, it could be suggested that first-person authority is only a matter of mastering a language. Adult persons who are competent language-users have learned a stock of everyday phenomenological vocabulary, and they are considered to be beyond criticism in their psychological self-ascriptions just by virtue of the fact that they generally use that vocabulary in a coherent and consistent manner, without regularly coming into conflicts with other competent language-users. Upholding the first-person authority might be seen as a mere pragmatic or social convention.

If we think this way, how unassailable a subject's descriptions of her conscious experiences are will be a function of her adeptness in using experience-vocabulary. The descriptions of a fully competent adult will be authoritative, the descriptions of a young child or a non-native speaker less so. However, what should we do in situations where two people, both perfectly competent in introspecting and describing conscious experiences and who we have independent reasons to believe to be undergoing a similar experience, nevertheless describe that experience in mutually inconsistent ways? Do we then have to assume that at least one of them makes an introspective error? Are we then entitled to waive the first-person authority of one or both of them? For Schwitzgebel (2011), cases like that form the basis of one group of arguments to the effect that people are not in general reliable judges of their own conscious experiences.

Constitutivism

If Schwitzgebel is right, much of the first-person authority that we normally grant to competent adult people is based on false prejudice. However, there is an alternative view of self-knowledge that denies that describing our experience is essentially a matter of having an accurate perception of one's inner episodes, which is then translated into words. This view, called "constitutivism" by Finkelstein (2003), is also friendlier to first-person authority than detectivism ends up being. Its central idea is that our judgments concerning our inner episodes play a constitutive role in determining what those inner episodes are.

Constitutivism seems insightful especially concerning propositional attitudes like beliefs. When we self-ascribe a belief, it seems that we most typically do

that by *rationally committing* ourselves to a belief, via judging that something is the case. Self-ascribing a belief seems to be the act of forming a belief or settling on a belief, rather than finding one via introspection. As the so-called transparency theories of self-knowledge have emphasized, self-ascriptions of attitudes need not involve any judgment turned inwards, so to speak; they are rather part and parcel with the judgments we make of the outside world.

So at least in some cases, my non-introspective judgments may *constitute* my mental states. Also in the case of descriptions of sensations, my authority may be thought to be "not like the authority of an eyewitness [. . . , but rather like] that of an Army colonel when he *declares* an area off limits" (Finkelstein 2003, p. 28; emphasis in the original). A slightly adapted example will illustrate the point:

Example 2

> Two people have an abdominal pain that they describe to a physician. It has been established that their pains are caused by a similar medical condition; they are of the same age, gender and build, the patterns of activation in their nervous systems are highly similar, and their pain-descriptions agree for the most part. In short, we have good independent grounds for believing that they are describing qualitatively similar experiences.
>
> One person describes her pain as sharp and ripping. The other person disagrees, saying: "I don't think it is ripping at all, not really sharp either. It's more like crushing and suffocating."

Maybe we always have to leave some room for the possibility that, despite all clues to the contrary, the subjective experiences of the two people are, after all, different. But even if we assume that the experiences are similar and the subjects are just giving mutually incompatible descriptions of the same pain, we can interpret this as a case of faultless disagreement.

We can suggest that what the subjects are doing is not that they observe by introspection features of their inner experiences accurately or inaccurately. Instead, they are making spontaneous applications of concepts, and in doing this they engage in *defining* what their experiences are like. They are flagging a certain description as the correct thing to say about their experience. First-person authority, according to this view, is a matter of being in the unique position of choosing how experience-vocabulary is to be applied to one's subjective experience. What ultimately makes it the case that a subject's pain is ripping is the fact that the subject *judges* it to be ripping. Even if there is another, incompatible description of a qualitatively identical experience — even if the description of the first subject is highly anomalous — there is no need to ascribe error to any party. The deviant description can be treated just as a different application of experience-vocabulary, an application that is within the subject's rationality to make, and which may be psychologically interesting in itself. It does not force us to waive the first-person authority of any speaker involved. First-person

authority is the acknowledgement that subjects' statements about their experiences are (treated as) true in their own conversational context.

Constitutivism, in the case of describing my pain, would be friendly to first-person authority by holding that my sincere testimony is the primary court of appeal which determines what my pains are like. The fact that I judge my pain to be ripping plays a constitutive role in making it the case that my pain *is* (rightly characterized as) ripping. First-person authority exists, according to this view, because the primary way of establishing the character of someone's experiences is to refer to that person's sincere avowals about those experiences. For that reason, my judgment to the effect that I have a ripping pain is essential in making it the case that my pain is indeed ripping, as opposed to crushing or suffocating. Of course, there will be constraints on how I can describe the pain; I cannot normally characterize my pain as "dark green" or "prestigious," for example. But it can be argued that this would not be because those descriptions are erroneous in light of some independent standard, but because they violate some conversational maxims; I would normally know that those words are probably uninformative to others as pain-descriptions. Insofar as I want to communicate, I should not use unhelpful concepts, but otherwise I am free to describe my pain in whatever way seems to me most suitable. In determining what is true to say about my experiences, those avowals of mine will be the primary point of reference. First-person authority just reflects this state of affairs.

Is constitutivism preferable to detectivism? Two points of criticism are important. First, it cannot really be praised as an account of self-knowledge. Instead, it makes it hard to characterize my pain-descriptions and other avowals as instances of (self-)*knowledge* at all. Knowledge conceptually requires some kind of systematic avoidance of error. Roughly speaking, if something counts as an instance of knowledge, it should involve a judgment that succeeds in representing some state of affairs *correctly*, in virtue of some laudable systematic method. If constitutivism generally holds, and truths about persons' inner states are primarily determined by referring to their avowals, then there will be no such thing as the cognitive achievement of getting a psychological self-description *right*. It will be no more of a cognitive achievement than launching an arrow into a wall and drawing a bulls-eye around its head is an archery achievement.[2]

Second, constitutivism seems to make us responsible for mental facts about ourselves in a way that is not plausible across the board. Here it becomes evident why constitutivism fits better together with accounts of beliefs and other similar attitudes. When we consider the latter, constitutivism seems advantageous, because we generally want to be personally responsible for the contents of our

[2]I believe that something like this thought is behind those remarks of Wittgenstein that suggest a "non-cognitive thesis of avowals," as Hacker (1975) calls it.

beliefs and desires. But sensations are different in this regard. According to constitutivism, what ultimately makes it right to say that my pain is ripping instead of crushing is the fact that I judge it to be ripping instead of crushing. But in many cases, I will be unable to accept this account from my own viewpoint. It will at least usually, if not always, strike me as false to say that my pain is ripping because I judged it to be ripping. In a typical situation, I say my pain is ripping because my pain calls for exactly that word, and I will be inclined to insist that I really have no rational control over that matter. If I complain of a sharp pain, no one can seriously suggest to me: "Learn to *judge* it to be dull instead, and then it will not be sharp anymore!" Not all conscious experiences, as they appear to me in first person, leave room to intellectually decide the most appropriate verbal characterization for them. Some experiences do not let me rationally judge what I want to say of them; they will rather take control of me, and demand an expression. This uneasiness from the first-person viewpoint should justify looking for a better account to surpass both detectivism and constitutivism.

Expressivism

Finkelstein (2003), Bar-On and Long (2001), Bar-On (2004), and Rodríguez (2012) have examined expressivism as a superior alternative for making sense of our relation to our own inner sphere. This view develops a point inherited from Wittgenstein (1953), saying that much of psychological talk in the first person is not descriptive in nature; it does not stem from an observation of an inner object. Instead of merely rejecting detectivism, however, Wittgenstein insisted on continuity or at least a possible connection between verbal avowals and primitive, "natural" expressions:

> How do words refer to sensations? [. . .] The question is the same as: how does a human being learn the meaning of the names of sensations? — of the word "pain" for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior. "So you are saying that the word 'pain' really means crying?" — On the contrary: the verbal expression of pain replaces crying and does not describe it. (Wittgenstein 1953, §244)

According to the possibility Wittgenstein points out, the avowals that we use to talk about our experiences work in the same way as pre-verbal grunts and cries. The point of the avowals is not to be parts of fact-stating discourse, but to give voice to wants and needs in social interaction. The avowals can also be drawn out of me against my will, like primitive expressions. This is a point in favor of expressivism against constitutivism, as the latter threatened to over-intellectualize the subjective sphere.

For the question of why my descriptions of my own experience carry a special authority, expressivism offers a deflationary answer. According to it, avowing is not a matter of describing one's pains or feelings at all. Avowals only superficially look like descriptions. Actually they are sophisticated and cultured expressive behavior: utterances that are in the business of reacting to my surroundings, and thereby doing other things, such as eliciting pity or asking for help. This was Gilbert Ryle's view in his *Concept of Mind*:

> [M]any unstudied utterances embody explicit interest phrases, or what I have elsewhere been calling "avowals," like "I want," "I hope," "I intend," "I dislike," "I am depressed," "I wonder," "I guess," and "I feel hungry"; and their grammar makes it tempting to misconstrue all the sentences in which they occur as self-descriptions. But in its primary employment "I want…" is not used to convey information, but to make a request or demand. [. . .] Nor, in their primary employment, are "I hate…" or "I intend…" used for the purpose of telling the hearer facts about the speaker; or else we should not be surprised to hear them uttered in the cool, informative tones of voice in which we say "he hates . . ." and "they intend . . . ." We expect them, on the contrary, to be spoken in a revolted and a resolute tone of voice respectively. (Ryle 1949, pp. 183–184)

However, even if Ryle's view of the primary employment of avowals is correct, he realizes that he cannot boldly generalize this point. The existence of a "primary" employment implies that there are one or more secondary employments. Surprising or not, sometimes "I hate . . ." and "I intend . . ." *are* uttered in a cool and measured manner, in order to give a self-description. The view that avowals are simply expressive and lack truth-values is rightly met with suspicion (Hacker 1975; see also Malcolm, 1954). Obviously, if this simple view is what expressivism amounts to, it will explain (apparent) first-person authority, but it will not be an account of self-knowledge. According to it, my verbal avowals are no more instances of self-knowledge than distinctive grunts and gestures are. On the other hand, those avowals cannot be meaningfully corrected by another person, but this is for the trivial reason that they have no factual content to disagree on.

Wittgenstein (1953, II, ix) plausibly acknowledged that avowals can play the role of both expressions and descriptions, or something in between. A non-naive version of expressivism holds that my speech about my own mental states is fundamentally continuous with my natural bodily expressions, but such speech still linguistically expresses or "manifests" facts about my thoughts and feelings so that it is capable of stating truths or falsehoods about me. Bar-On (2004) has developed such a version and labeled it "neo-expressivism." Sophistication is clearly necessary, because it is hard to deny that avowals are in some sense also in the business of stating facts about their speaker. Avowals have contents that can feature in logical inferences, they can be contradicted by other statements, and so on. It seems that expressivism has to face an objection that is parallel to the Frege–Geach problem for metaethical non-cognitivism (for a summary, see Sinclair, 2009): How can this way of talking be fundamentally expressive, when it evidently in many contexts functions like descriptive, fact-stating talk?

In what follows, I will present a development of expressivism to shed light on the nature of avowals, the first-person authority associated with them, and the limitations of that authority. I attempt to combine a number of what I take to be valid insights. First, I will endorse a view that I attribute to Wittgenstein: avowals can function as expressive utterances but also as descriptions, and there is no categorical line separating the two cases. Second, I agree with Rodríguez (2012) in holding that Bar-On's (2004) influential expressivist account has the undesirable feature of taking apart avowals as expressive *acts* and avowals as the linguistic (truth-evaluable) *products* of those acts. I suggest that the putting forward of a linguistic description of one's experience is a single expressive act, whose expressive quality and truth-value are assessed in an interdependent fashion. Third, I seek to integrate detectivism and constitutivism in the picture, by highlighting the kinds of cases where each works best.

Primitive Avowals, Intellectual Self-Ascriptions, and Deliberations

For heuristic purposes, I will distinguish between three different types of psychological self-ascription. These are not meant as rigid categories. Instead, they represent the end and middle points of a scale on which avowals, and interpretations of avowals, can move. One extreme is a purely expressive, spontaneous avowal; another extreme is a detached, cool self-ascription done as if from a third-person perspective. Between these, there is a vast range of avowals that express the speaker's state of mind *by* asserting something about it. A good label for these latter cases is hard to come by; I will call them *deliberations*, owing the word, and some of my inspiration, to Moran (2001).

Other advocates of the expressivist view have made the point that (some) avowals have a special epistemic authority because of their peculiar expressiveness. They are taken to be immediate, non-judgment-involving airings of the subject's mental states. My aim is to qualify, and clarify, this point by suggesting that some avowals (deliberations) have a special epistemic authority when they are expressive in a certain spontaneous and unstudied way while *also* being honest attempts of a revisable self-description.

Primitive Avowals

First, I endorse Wittgenstein's point about verbal expressions of feelings being able to take over and extend the function of primitive, non-verbal expressions. Assuming that more articulate and considered expressions can build on simple primitive expressions, I will propose a way of seeing these as a procession on a single, continuous scale. Primitive, natural expressions like cries and smiles are devoid of cognitive content. They are not attempts to convey factual information. They may be expressions of attitudes, means of drawing attention, devices of eliciting reactions from others or otherwise communicative, but they are not

statements or descriptions of the subject's mentality. They can be called purely expressive acts. The simplest form of verbal avowals can be equated with them. Cases where "It hurts!" is used spontaneously and passionately to serve the same function as would be served by a scream, or a case where a spontaneous "I feel so good!" takes the same communicative role that could be taken by an exhilarated smile, can be called purely expressive avowals. These have a character of naturalness and spontaneity; they are drawn out of a person, rather than formulated and put forward by the subject in a controlled fashion. This is one end of my proposed spectrum.

Intellectual Self-ascriptions

On the other end of the spectrum, there are self-ascriptions of mental states that are purely descriptive. Whereas purely expressive avowals are not descriptive to any extent, the self-ascriptions of mental states at the other end of the spectrum are not expressive to any extent. The latter are instances where the subject takes a detached, third-person perspective toward her own mentality, and produces a studied verdict from that perspective. She may or may not like the contents of that verdict; she may even want to disown it. I will call these intellectual self-ascriptions. They will include a case where I reluctantly admit, after a lengthy work to sort out my thoughts, that I am angry with my father because of his strictness as a disciplinarian, while at the same time admitting that I should not and do not want to be angry with him. In another case, I notice my slowing pace of work and carelessness and conclude that I must be tired and frustrated, although I do not feel like saying that I am either of those things; but my physical and behavioral condition force me to make that conclusion anyway. I know, after all, that lethargy and carelessness are objective criteria for a person's being tired.

At this latter end of the spectrum, it can be legitimately said that I come to know my own mental states by *detecting* them in myself, although that detection is not necessarily carried out by inward glances of introspection. In any case, in these instances I attribute a mental state to myself as a result of self-observation of some kind, and this observation has no special claim of authority over anyone else's word. My self-observation can be mistaken for the same mundane reasons as any observation can be mistaken; it will make perfect sense to ask me to do my self-observation more carefully or more attentively, in order to avoid error. It is possible that I mistake the symptoms of a medical condition for symptoms of tiredness, or that I misidentify as repressed anger something that further reflective work reveals to be some other complex feeling. In short, this is a class of cases where I am sufficiently alienated from my own mental state to treat that mental state as an external object of scrutiny. The account of detectivism, while not easily generalizable, fits well here. This kind of self-scrutiny was what

Ryle (1949), who rejected privileged introspective access as the basis of self-knowledge, eventually treated as the paradigm case of real self-knowledge.

Deliberations

There is a purely expressive case of avowal; these I have called primitive avowals. There is also a purely descriptive case of avowal (according to my liberal use of the term); these I have called intellectual self-ascriptions. Now I will distinguish a third case, which is *the speech act whereby the subject puts forward an expressive linguistic utterance to serve as a self-description*. I believe that many, maybe most, avowals in typical human communication can be seen as instances of this type. They are characterized by a desire of the speaker to strike a balance between saying something that can be taken to be an objectively accurate description of her, and voicing her own impulses and wants, all in a single speech act. They are expressive utterances of the subject, but these expressive utterances acknowledge that they are attempts at manifesting a mental event that is an object of scrutiny also from the subject's own perspective; an event for which the giving of an adequate description is a cognitive challenge. They are instances where the subject assesses two things at once: first, what she wants to say about her experience; and second, how objectively plausible her statement is as a self-description. I call these avowals *deliberations*. One more modified example will serve to illustrate the point.

Example 3

> I have an abdominal pain that I need to describe to a physician. I am asked to assess my pain's intensity on a scale from 1 to 10. I have used the pain-scale before, and I consider the guidelines I associate with different numeric degrees of pain. I judge that my pain is of the level of 7. Then I am asked to think carefully:
>
> "You describe your pain otherwise in the same way as in those earlier instances when you have judged it to be 5 or 6. You also don't show signs of greater distress over it. Are you sure that 7 is not too much?"
>
> I answer: "Yes, I understand that, but I just feel that this week it is harder to bear. I'm not sure if it is the pain itself that intensifies or if I am just depressed, but 6 would be too small a number now. I'm saying 7."

Here, I am doing several things at once. First, I am giving a description of my pain. My utterance of "7" occurs in a context of giving a description; it is meant to inform the other about a certain feature of my conscious experience, to go down in my medical record as a true proposition about my condition. Second, I am using words (or rather, numbers) expressively: the point of my saying "7" is to let the other know how I feel about my pain, to voice my sentiment. Third, however, in this particular example I acknowledge that I have

some reservations about whether my avowal accurately describes a change in the pain itself or in the overall quality of my mental condition ("I'm not sure if it is the pain itself that intensifies or if I am just depressed"). Here I admit that my decision to say 7 instead of 6 might be borne out of my growing concern over my pain, my overall feeling bad physically and mentally, or something like that. In a way, I give the hearer some freedom to evaluate what conclusions to draw from my utterance.

Now, it seems to me plausible to agree with constitutivism to an extent. My description of my pain has a unique claim to being true. This is because my avowal has a special status in determining what is deemed right to say about my pain's intensity. My honest avowal of my pain as 7 is a central criterion for it being the case that my pain indeed has the intensity of 7. I am the only one who can apply pain-vocabulary to myself in the first person, so my judgments about my pain are crucial in determining how pain-vocabulary is to be applied to me in particular cases. However, my avowals are not the only criterion for determining what my pain is like; there are bodily and behavioral criteria for different kinds of pains too (as the interlocutor in Example 3 notices).

In light of this, I suggest that my avowal is a complex communicative act: it is, in effect, *a request for others to accept my pain-avowal as a valid description of my pain*. It has a double nature. It is put forward as a description of my state of consciousness, but it is also a kind of an act of pleading: an expression of my want to make others treat my pain as a pain of the level 7. Most of the time, my avowal will be accepted as a valid and authoritative description without a scruple, insofar as people generally accept the first-person authority of subjects over their own mental states. But sometimes there will be room for scrutiny, as in Example 3.

In Example 3, I am saying that my pain is 7 in circumstances where, as far as any onlooker can see, I could as well say 5 or 6. So why am I saying 7? If this unusual question would be put to me, I could approach it in a number of ways.

(a) First, I could try to ground my judgment in some objectively available behavioral evidence. "Look, I may not show signs of greater distress over my pain just now, but there are some signs anyway: it distracts me more than before, it is harder to concentrate on anything else, I am constantly more stressed about it than before . . . . It must have intensified from 6 to 7." Here, I am taking a more detached position toward my pain by allowing that it is a matter of evidence to decide whether my pain is 6 or 7. This is to move my avowal more in the direction of what I have called an intellectual self-ascription. I loosen my claim to first-person authority somewhat, by allowing that my judgment about my pain might be wrong according to some standards that can override my own statement.

(b) Second, I could (in principle) decide to be a hard-headed constitutivist. "I just feel like saying 7. It seems to me to be the correct application of the

pain-scale to what I am feeling right now. And I am automatically right in this, because it is me who gets to decide how pain-vocabulary is applied to my inner experiences. End of story." I think it is evident that by these words, the subject would make her avowal sound in a certain way suspicious. It seems that her attitude toward her pain is not an attitude of a person who wants to communicate something about her pain to others. It is rather the attitude of a person who is merely interested, for one reason or another, to ensure that the hearers withhold further inquiry and accept her statement. Concerns about the honesty of the avowal would be raised, and there would be some hesitation about whether, or to what extent, her utterance can be taken seriously as an avowal. It would be sensible to protest that the subject does not get to *decide* whether his pain is 6 or 7 just like that.

(c) Finally, there is the option that seems to be natural and plausible: "I just feel that I have to say 7. I cannot help it. It just feels worse today." What I acknowledge here is that my avowal shares the nature of a primitive expression: the number 7 is drawn out from me, somewhat in the way spontaneous grunts or smiles are drawn out from me, rather than rationally decided to be my chosen number for the pain.

Now, I suggest that the first-person authority of my self-ascription is at its strongest when it has a nature like that described in (c). When it is in this way akin to a spontaneous, primitive expression, then the subject has a special force behind her request that her pain-avowal is treated as a valid description of her pain. Her avowal will then represent her genuinely best effort to give a linguistic expression to an event of her consciousness that does not allow for just any arbitrary expression.

In other words, I am suggesting that the authority of an avowal as a self-description is dependent on whether the avowal is taken to share in the nature of a primitive avowal. But I am also arguing that it is necessary for an avowal to be plausible as a description from a detached perspective, if it is to work in its role as an avowal. Once more, I will illustrate by an example. Let us imagine that, in Example 3, I am struck with a sudden fear and anguish over my pain, and I start to feel my constant, familiar pain as so unbearable that I want help with it immediately at any cost. Then, when asked about the intensity of my pain, I will respond "10." Now, it seems that another person would have a good reason to say to me: "Look, you cannot really say that. I know you feel bad, but 10 is the highest point of the scale, it is meant to represent a pain that is so unmanageable that you have never experienced anything worse than it. A person with a pain that has the intensity of 10 would be incapacitated, which you clearly are not." In a way, my "10" would be a failed avowal; it could not be taken seriously as an avowal.

In the previous case, I am uttering "10" as a kind of a purely expressive call for help that does not even purport to be a measured attempt of self-description.

This kind of an avowal will be appropriate in some conversational contexts, but faulty in many others. In particular, it will be unhelpful for the physician, or at least it will put the physician in a position where she has to contemplate how to interpret my utterance. It will not be a fully functional avowal in its context.

I take these considerations to show the following. Insofar as my (deliberative) avowals are descriptions of my mental state, they are also requests for others to accept my description as valid. But the acceptability of my avowal as a valid self-description is largely dependent on whether my avowal is taken to be expressive in the right way (i.e., in the way of a primitive, unstudied expression). And my avowal, however honestly expressive, will not be fully taken seriously as an avowal unless it at least attempts to be a descriptive act (i.e., is constrained by my aspiration to inform others about what my pain is like, and not only by what I want to say about it). The descriptive and expressive aspects of an avowal are interdependent.

Conclusion: Avowals, Self-Knowledge, and the Nature of First-Person Authority

I will now conclude by spelling out some consequences for the issues of self-knowledge and first-person authority that can be drawn from my discussion. First, it seems to me that a crucial part of what is commonly called "self-knowledge" is manifested in a person's ability to reflect on her use of the different modes of avowals, and to some extent choose between them. Avowals are called for in many different communicative situations. Sometimes, when another person asks me "How do you feel?," what is expected from me is just a spontaneous manifestation of my feeling of pain, affection, or anxiety. Then, it is an exercise of self-knowledge to be able to recognize and let out my spontaneous and unstudied reaction, suppressing any need to take a detached perspective and survey my state of mind as a part of my objective personal psychology. At other times, it will be necessary for me to study my psychology as if from a third-person perspective, in order to uncover biases or unconscious motivations, acknowledging that my own assessment of my psychology is nothing but an assessment by a fallible human being. Then, it will be necessary to contain my spontaneous and unstudied reactions, and to keep in mind the possibility that my first thoughts about my pains, affections or anxieties might not be the (whole) truth about them. ("I feel like saying that this pain is 7; but don't I usually have a low pain threshold? Maybe most other people would call it 6, or even 5? And I admit that I am feeling depressed; maybe that is affecting all my judgments more than I realize.") Understanding that my unstudied expressions and correct descriptions of my psychology (according to some standards that I myself can accept when speaking in third person) can come apart, and finding out how they can be expected to come apart in diverse situations, is a vital part of my

self-knowledge. In deliberations, I talk expressively, and in so doing I manifest my wants and needs to characterize my mental life in certain ways, but at the same time I am subjecting my avowals to interpersonal assessment by presenting them as *descriptions* of myself. Seeing how those expressively grounded descriptions manage with and against those descriptions of me that are given from the perspective of another person helps me to cultivate an important kind of self-knowledge. I am learning how my conception of myself plays together with other people's conception of me.

This characterizes self-knowledge in a sense in which it is a process. It is a sense of self-knowledge that is easily overlooked if the crucial expressive function of first-person psychological talk goes unnoticed. In deliberations, how I can plausibly describe myself constrains how I should feel appropriate to express myself, and at the same time how I need to express myself constrains how I describe myself. Competent use of deliberative avowals might be characterized as communication that is at the same time both self-studying and self-defining — a remarkable feat of human thought.

Second, pointing out the combination of expressiveness and descriptiveness in avowals produces a modest and commonsensical view of first-person authority and its limitations. There is little motivation to assume that individuals have magically accurate introspective powers, so that they would be uniquely authoritative judges of their own mental states in a detectivist manner. But what people do have is a subject's perspective to those mental states, and a desire to define and characterize those states from that perspective. Conscious attempts of persons to work out what their subjective experiences are like — what I have called deliberative avowals — have a special epistemic status insofar as they are properly expressive honest utterances while also being attempts of self-description. A description that I give of my own experience is authoritative when, and insofar, it is based on an expressive act that is ungrounded and natural in the same way as a primitive bodily reaction is. The subject is the only one who is in a position to give a description with this peculiar basis; therefore, naturally, an avowal of this kind carries special weight. When moving away from deliberations toward primitive avowals, or toward intellectual self-ascriptions, motivation to demand a special authority for the avowals wanes: in the case of primitive avowals, because they are not issued or interpreted as statements with factual content, and in the case of intellectual self-ascriptions, because they are not made from the special perspective of the subject-position.

First-person authority is, first and foremost, recognition that each person has a unique status as a generator of knowledge about her own mental reality. Properly expressive deliberative avowals have a special epistemic job to do. They are not infallible, not always even highly reliable, but they are acts of giving voice to a personal experience: they are the subject's applications of concepts to her personal experiences in a certain situation and at a certain time, and as

such they have a constitutive role. They serve as the starting point of inquiry into her experience, and enjoy a certain amount of resistance to corrections. The role of such avowals as (partly) self-defining acts also means that a subject can, in principle, decide to stick to her self-description even when it is anomalous from the perspective of an outside observer. If a subject is truly brought to see her self-description as erroneous, this must happen by eventually bringing her to revise her avowal in such a way that she can, after the revision, own it as her honest self-expression, not only as a third-person description of her forcibly given from outside. This seems an essential characteristic of an autonomous, self-standing subject. Consequently, first-person authority has an ethical dimension in addition to an epistemic one. Respecting it is to grant to other people an authoritative voice in telling what their experiences are like. Disregarding it is to say that it is in principle possible to overrule a subject's self-expressing voice by a third-person, more authoritative account of what her experiences are *really* like. It is doubtful whether those who are subjected to the latter treatment have a chance of seeing themselves as subjects in the full sense.

References

Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Oxford: Oxford University Press.

Bar-On, D., and Long, D.C. (2001). Avowals and first-person privilege. *Philosophy and Phenomenological Research, 62,* 311–335.

Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.

Finkelstein, D.H. (2003). *Expression and the inner.* Cambridge, Massachusetts: Harvard University Press.

Gertler, B. (2011). *Self-knowledge*. London: Routledge.

Hacker, P.M.S. (1975). *Insight and illusion: Wittgenstein on philosophy and the metaphysics of experience*. London: Oxford University Press.

Hurlburt, R.T., and Schwitzgebel, E. (2007). *Describing inner experience? Proponent meets skeptic*. Cambridge, Massachusetts: MIT Press.

Malcolm, N. (1954). Wittgenstein's Philosophical Investigations. *The Philosophical Review, 63,* 530–559.

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton, New Jersey: Princeton University Press.

Rodríguez, Á.G. (2012). How to be an expressivist about avowals today. *Nordic Wittgenstein Review, 1,* 81–101.

Ryle, G. (1949). *The concept of mind*. New York: Barnes & Noble.

Schwitzgebel, E. (2011). *Perplexities of consciousness*. Cambridge, Massachusetts: MIT Press.

Sinclair, N. (2009). Recent work in expressivism. *Analysis, 69,* 136–147.

Wittgenstein, L. (1953). *Philosophical investigations* (third edition). [G.E.M. Anscombe, Trans.]. Oxford: Blackwell.

The Journal of Mind and Behavior
Summer 2014, Volume 35, Number 3
Pages 167–184
ISSN 0271–0137

"Feeling what Happens": Full Correspondence and the Placebo Effect

André LeBlanc

John Abbott College and *Concordia University*

This paper proposes a theory whereby the physiological changes induced by placebos are accompanied by corresponding changes in the patient's mental state. I begin by defining the placebo problem, and review the three leading theoretical approaches for solving it — meaning theory, expectancy theory, and conditioning theory — before discussing the significant theoretical issue posed by a classic case of placebo immunosuppression in rats. The theory of full correspondence is then introduced as a way of explaining the nature of the placebo effect and of resolving the conflict between "meaning-oriented" and "mechanism-oriented" approaches to the phenomenon. After proposing how to test the theory experimentally and examining existing evidence for it, I consider its ability to integrate the dominant theoretical perspectives of the placebo effect within a framework centered on the patient's subjective experience, the one variable overlooked on both sides of the meaning/mechanism divide.

Keywords: placebo effect, full correspondence, consciousness

All instances of the placebo effect seem to share the following feature: our ability to influence our bodies in ways that go beyond what is usually deemed possible. That we are capable of controlling our bodies, such as when we pour a glass of water, type an email, or hug a child, is not in itself unusual. What is unusual about the placebo effect, however, is our ability to influence bodily functions over which we do not normally have control, such as the neuronal activity of pain, the quantity of white blood cells in the immune system, or the brain chemistry of Parkinson's disease. The problem, then, is to explain how we exercise some measure of control over these apparently involuntary functions.

This research was supported in part by a Hannah Post-doctoral Fellowship in 2002–2003 (Department of the History of Science, Harvard University) from Associated Medical Services, Inc., Toronto, Ontario, Canada. Correspondence for this article should be addressed to André LeBlanc, Department of History, Economics and Political Science, John Abbott College, 21,275 Lakeshore Rd., Ste. Anne de Bellevue, Quebec, H9X 3L9 Canada. Email: andre.leblanc@johnabbott.qc.ca

Placebo theorists have been attempting to solve this problem for some time. They have generally come down on one side or the other of the mind–body divide: on the mind side, "meaning-oriented" researchers have focused on the social and psychological aspects of the placebo effect, while on the body side, "mechanism-oriented" researchers have concentrated on the placebo's physiological features (Harrington, 1997). To borrow a simile from the great German physiologist, Ewald Hering (1834–1918), these orientations are like two teams of engineers digging from opposite sides of a mountain and trying to meet at some point in-between (Turner, 1994). There is a general belief, moreover, that the joining of these two tunnels would enable us to solve the mystery of the placebo effect. That is to say, any satisfactory solution to the placebo problem, as defined above, would have to satisfy the additional requirement of overcoming the epistemological barriers that separate these diametrically opposed schools of thought.

This paper proposes to fulfill both these requirements in a theory based on the following supposition: feelings accompany placebo effects in the same way that feelings such as embarrassment, hunger, or sexual arousal also exist alongside their corresponding physiological effects. In other words, I shall argue that the nature of the placebo effect is easily explained if we recognize the possibility that the physiological changes induced by placebos are accompanied by corresponding subjective experiences. That there should be a specific brain state for each of our mental states is virtually a universally accepted assumption among brain/mind theorists. Indeed, with each new advancement in brain research, we expect to find "ever finer correspondences between brain states and mental states, between brain and mind" (Damasio, 2002, p. 8). What I am proposing is the extension of this correspondence to mental functions whose existence we have yet to consider, namely, those associated with the neuronal activity of placebo effects.

Other theorists (Benedetti, 2009; Kirsch, 1997) have recognized a correspondence between the mind and body in placebo effects, but not to the extent considered here.[1] For example, Kirsch (1997) distinguished between two types of physiological responses to placebos, which I will hitherto call type I and type II

[1]There is one exception, however. In 1869, the Belgian philosopher, mathematician, and psychologist Joseph Delboeuf (1831–1896) proposed a similar idea to explain the case of Louise Lateau, a famous Belgian stigmatic. In April of 1868, still weak after recovering from a near fatal illness, this 18-year-old woman began bleeding from her left side, feet, hands, and forehead over a series of Fridays shortly after Easter. During these bouts of stigmata, Lateau was actively engaged in imagining the final moments of the passion of Christ. Given the close match between her bodily lesions and the contents of her overexcited imagination, Delboeuf (1869/1993) ventured the following hypothesis: "In certain exceptional and morbid cases, could not the felt sensation be joined by the corresponding organic modification [. . .]?" (p. 400). Delboeuf would go on to develop similar ideas, but in a slightly different direction, when he took up the study of hypnosis some 15 years later. It is only after I had hit upon the theory of full correspondence that I realized Delboeuf had already proposed a similar theory in 1869.

physiological placebo effects. Type I physiological placebo effects are assumed to come in mind–body pairs: the subjective experience being in close correspondence with its physiological counterpart. Examples include the psychological and physiological effects of placebo coffee and placebo tranquilizers. The physiological responses of type II placebo effects, on the other hand, "are not part of the physiological substrates of subjective experience" (p. 179); they have no counterparts in the mind. Kirsch points to the influence of placebos on cancer, skin conditions, and the immune system as examples of this more mysterious type of placebo effect. In sum, one could say Kirsch subscribes to a *partial* correspondence between mental and physiological events in the placebo effect, whereas I am proposing a *full* correspondence between the two.

If the idea of full correspondence has not been seriously considered until now (aside from Delboeuf's [1869/1993] and Kirsch's [1997] considerations), it is because we have paid insufficient attention to the feelings associated with the placebo effect. The reason for this oversight, as I shall later discuss, is tied to a deep-seated reluctance — similar to the skepticism of many researchers and theorists (e.g., Hróbjartsson and Gøtzsche, 2001; Kienle and Kiene, 1997) regarding the reality of the placebo effect — in recognizing consciousness as an acceptable object of scientific investigation. To many, consciousness and the placebo effect seem less real than the material objects that lie at the basis of our scientific understanding of the natural world. But to deny the existence of the placebo effect is to deny a well-documented natural phenomenon (e.g., Benedetti, 2009; Harrington, 1997; Moerman, 2002b), and to ignore consciousness in the study of that phenomenon is to ignore a vital clue in understanding it.

This paper will describe the theory of full correspondence in some detail, present evidence supporting it, and discuss its capacity to integrate existing theories within a single theoretical framework. My first step will be to situate the theory of full correspondence within the field of placebo research by reviewing its dominant theories — meaning theory, expectancy theory, and conditioning theory. This brief review is modeled on Anne Harrington's (1997) classic review of the placebo literature, which first introduced me to the epistemological tension described above and set the context for the problem addressed herein.

Meaning Theory

Building on the work of scholars sensitive to the role of culture and meaning in the placebo effect (Brody, 1997; Hahn, 1985, 1995; Kleinman, 1986, 1998), Daniel Moerman's (2002a, 2002b) "meaning response" theory, recently revised and expanded by Barrett et al. (2006) and Kradin (2004), draws on a rich history of studies revealing the symbolic and cultural factors involved in the placebo effect. Such studies have shown, for example, that two placebo tablets work better than one (Rickels, Hesbacher, Weise, Gray, and Feldman, 1970), that

capsules work better than tablets (Hussain and Ahad, 1970), injections better than pills (de Craen, Tijssen, de Gens, and Kleijnen, 2000), branded better than unbranded pills (Branthwaite and Cooper, 1981), and expensive better than inexpensive pills (Waber, Shiv, Carmon, and Ariely, 2008). In such pairs, Moerman (2002a, 2002b) noted, the former "means" more than the latter. Meaning theorists have drawn similar conclusions from studies showing how warm colors (pink, orange, and red) are consistently associated with stimulants and cool colors (green and blue) with sedatives and depressants (de Craen, Roos, Leonard de Vries, and Kleijen, 1996), how Chinese Americans born in unlucky years according to Chinese astrology tend to die younger than cohorts born under luckier stars (Phillips, Ruth, and Wagner, 1993), and how a warm, enthusiastic, and caring bedside manner increases the overall effectiveness of treatments and placebos (Di Blasi, Harkness, Ernst, Georgiou, and Kleijnen, 2001).

What these and countless other studies show is that our biology is deeply affected not only by the material basis of life, but also by the broader social world. As Barrett et al. (2006) wrote with respect to coffee drinking in particular, and "health-related behaviors" in general, we are "embedded within sociocultural networks of meaning. Conscious and subconscious 'meanings' combine with personal experiences — physiological and psychological — to form mind–body response patterns" (p. 189). For similar reasons, Moerman (2002a, 2002b) sees the placebo effect as a special case of a larger biosocial phenomenon he calls the "meaning response."

But meaning cannot tell the whole placebo story. We need to explain how "sociocultural networks of meaning" translate into physiological effects, how meaning moves from society to the body. One way of doing so is through the study and manipulation of expectations.

Expectancy Theory

Expectancy theorists take a more psychological approach to the placebo effect. They do not dispute the ideas advanced by meaning theorists, but they consider sociocultural factors to be a step removed from the psychological processes that produce the phenomenon, and the most important of these processes, they argue, is expectation. On this view, the placebo effect occurs when a patient is led to believe the treatment will have the desired effect. A classic example is an experiment using "trivaricane," a name invented by Montgomery and Kirsch (1996) for a placebo anesthetic cream they used to lessen the pain of unpleasant electrical stimulation. While the electric shocks were administered to both index fingers of their undergraduate participants, the placebo cream was applied to only one of the fingers. To enhance the placebo's effect, the researchers wore white lab coats, drew the cream from a bottle labeled "Trivaricane: Approved for research purposes only," and applied it wearing surgical

gloves "to avoid overexposure" (p. 175). The expectations created by this trivaricane-charged context produced a significant anesthetic effect, and because this effect was limited to the finger on which the cream was applied, appeals to general pain-relieving processes such as endorphin release or anxiety reduction were effectively ruled out.[2] Similar studies of placebo alcohol (Hull and Bond, 1986), placebo (decaffeinated) coffee (Flaten, Aasli, and Blumenthal, 2003; Kirsch and Weixel, 1988), anti-depressants (Kirsch and Sapirstein, 1999), and sedatives (Jensen and Karoly, 1991) have also provided strong support for expectancy theory.

Conditioning Theory

The classical conditioning theory of the placebo effect has a lot of experimental support too, with studies demonstrating behaviorally conditioned effects ranging from the reduction of pain, depression, and anxiety to the production of antibodies, insulin, and dopamine (Benedetti, 2009). Unlike the psychosocial orientation of meaning and expectancy theories, conditioning theory interprets the placebo effect as a type of associative learning. According to the conditioning account, for example, placebo aspirin works by inducing the pain relief previously associated with aspirin pills. In the language of classical conditioning, an aspirin pill is an unconditioned stimulus that produces the unconditioned response of pain relief. As the stimuli associated with aspirin pills, such as their taste, shape, and color, are repeatedly paired with the unconditioned stimulus, they become conditioned stimuli capable of triggering conditioned responses similar to the unconditioned response produced by the pills' pharmacological agent. In short, all the stimuli that had previously been associated with a medical treatment have the potential of eliciting that treatment's physiological effects when a placebo is substituted in its place (Ader, 1997).

The conditioning approach does not deny a role for meaning and expectation in many placebo effects, but it often sees this role as secondary to the primary one of conditioning because, as we shall see below, some conditioned placebo effects seem to occur in the absence of any conscious cognition. Meaning and expectancy theorists naturally take the opposite view, subsuming conditioning within their own explanatory frameworks whenever possible. For example, Kirsch (2004) believes "expectancy theory includes conditioning as a process by which expectancies are formed" (p. 341). Under certain conditions, moreover, the placebo effect will correspond to a subject's expectations even when conditioning predicts the opposite outcome (Kirsch, Lynn, and Miller, 2004; Montgomery

[2]Some of the subjects could have inferred that the anesthesia would extend to the other index finger because the cream had been absorbed into the bloodstream. None of the subjects seemed to have formed this expectation, however, since the effect was limited to the finger upon which the placebo cream was applied.

and Kirsch, 1997). And while acknowledging the apparent absence of expectancies in Benedetti, Amanzio, Baldi, Casadio, and Maggi (1999), discussed below, Kirsch (2004) nonetheless maintains that "conditioned placebo effects without expectancies are rare" (p. 342). Yet much turns on these rare cases, the most famous of which was a chance discovery by Robert Ader in 1974 of a conditioned immune system response in rats.

Meaning versus Mechanism: Interpreting a Placebo Effect in Rats

Ader (1974) had initially set out to determine whether rats could be made to avoid saccharin-flavored water by inducing an association between the taste of saccharin and the experience of nausea. He began by giving groups of rats 1, 5, or 10 ml of water containing 0.1% saccharin, followed 30 minutes later by the injection of a nausea-inducing drug called cyclophosphamide, which also happens to be a powerful immunosuppressor. Control groups received the saccharin solution without cyclophosphamide. The rats were offered the same saccharin solution every three days and, as expected, the degree to which they avoided the flavored drink was found to vary with the quantity of saccharin consumed on the day of conditioning. Near the end of the experiment, Ader noticed something unexpected: several rats in the cyclophosphamide groups began dying, despite having received doses well below toxic levels. Moreover, as a general rule, the first rats to die received the largest volume of saccharin water in the initial pairing, the next rats to die, the second largest, and so on and so forth. Ader thus hypothesized that the rats had been conditioned to suppress their immune systems whenever they drank saccharin-flavored water, thereby leaving them vulnerable to pathogens in their environment. To test this hypothesis, he and a colleague subjected rats to a similar procedure in a subsequent study (Ader and Cohen, 1975), except this time they also injected the rats with sheep's blood and measured the quantity of antibodies produced by their immune systems. The results were as they had predicted. After a single pairing with cyclophosphamide, the saccharin alone acted as an immunosuppressor. These rats had been conditioned to respond to saccharin as if it were cyclophosphamide, just as Pavlov had conditioned his dogs to salivate at the sound of a bell after associating that sound with the arrival of food.

This experiment therefore showed that the placebo effect can be the result of processes that appear to be entirely mechanical, and because the effect was so clearly automatic and quantitative, it also suggested the possibility of explaining the fundamental mechanism of the placebo effect without recourse to expectancy, meaning, or any other cognitively based theories.[3] To reconcile these theories with conditioning, we might be tempted to apply expectancy

[3]As Harrington (1997) put it, "[t]he fact that [Ader] had achieved [a physiological placebo effect] in rats rather than in humans [. . .], undermined the frequent assumption that placebo effects were the product of peculiarly human interpersonal processes and unconscious wishes" (p. 6).

theory to Ader and Cohen's rats. After all, rats are surely capable of cognitions as commonplace as expectation. There are reasons to believe Ader and Cohen's experiment would not support expectancy theory, however, even if it were carried out with human beings. Stewart–Williams and Podd (2004) provided an elegant argument to this effect in their review of the expectancy-versus-conditioning debate over the mechanism of the placebo effect. They illustrated their argument with a personal example by M.E.P. Seligman, who, having caught the flu several hours after eating a meal with Béarnaise sauce, was later surprised to discover that the mere thought of tasting his favorite sauce produced strong feelings of nausea. If Seligman was surprised by his discovery, it is because the nausea came upon him unexpectedly, which is not what expectancy theory predicts. By analogy, it seems likely most conditioned taste aversion experiments operate in the same mechanical way and are largely oblivious to what the subject's expectations, hopes, or beliefs may be. Moreover, although the broader question of whether conditioning can occur in the absence of awareness is "long-standing and vexed," Ader and Cohen's experiment, along with other similar findings (e.g., Benedetti et al., 1999), seem to provide "persuasive evidence that conditioning in humans is not always cognitively mediated" (Stewart–Williams and Podd, 2004, pp. 332–333). Speaking specifically to this point, Benedetti (2009) added "there is experimental evidence in humans that *unconscious conditioned placebo responses* [emphasis added] are present in the immune and endocrine system (chapter 6) and in the cardiovascular and respiratory system (chapter 7)" (p. 45).

Unsurprisingly, Moerman (2002b) is uncomfortable with the Ader and Cohen (1975) study. He sees it, along with other conditioning experiments (for example, Benedetti et al., 1999, discussed below), as illustrating some of the limitations of his meaning response theory. When Pavlov's dogs learned to salivate at the sound of a bell that had previously been associated with food, Moerman (2002b) assumes "that the dogs didn't 'know' that the bell 'meant' food, that is, that their reactions were not cognitive ones involving understanding or meaning" (p. 124). Moerman is thus forced to concede instances of the placebo effect in animals, and possibly humans, that culturally oriented approaches seem powerless to explain.

The Theory of Full Correspondence

We are thus faced with two fundamentally different theoretical approaches to the placebo effect: one that explains the phenomenon in terms of meaning and expectations, and the other that explains it in terms of conditioning. My task will be to subsume these two approaches under a more general one. As I shall now argue, this more general approach to the placebo effect consists of a comprehensive correspondence between the subjects' mental states, on the one hand, and their physiological states, on the other.

With respect to Ader and Cohen's (1975) rats, the two main theoretical approaches to the placebo effect are easily reconcilable if we accept the following proposition: each time the rats drank the saccharin water after the initial pairing, the taste triggered the *memory* of how they felt the first time they tasted the sugary solution. In other words, on re-tasting the saccharin, the rats were reproducing not only the physiological effects of the cyclophosphamide, but the corresponding *psychological* effects as well. The initial injection of the toxin made the rats feel sick, and sick in a particular way, and that particular feeling was recalled each time they tasted the artificial sweetener. The more saccharin they tasted in the initial pairing, moreover, the stronger the subsequent association between the taste of the saccharin, on the one hand, and the biological *and* psychological effects of the cyclophosphamide, on the other. Thus, if the theory of full correspondence is true and the rats were remembering the feeling produced by the cyclophosphamide, "meaning oriented" and "mechanism oriented" approaches to the placebo effect can now meet on the common epistemological (and ontological) ground obtained by "*the feeling of what happens*" when the placebo effect occurs.

As mentioned before, theorists already widely assume that for each mental state there exists a corresponding state of the brain. The theory of full correspondence extends this assumption to mental states whose existences have been hitherto overlooked and whose neurological correlates include, but are not limited to, all type II physiological placebo effects. Full correspondence thus views the nature of the placebo effect not so much in the mechanisms by which it is produced, as in the correspondence between the subject's mental and physical states *when* it is produced, regardless of the mechanism at work. The cues that trigger the placebo effect need not even be conscious, as demonstrated by Jensen et al. (2012); but at the moment the placebo effect occurs, full correspondence predicts that the observed physiological modification (or an earlier physiological trigger that led to this modification)[4] will be accompanied by a matching psychological modification. Returning to the problem with which I began my inquiry, the secret, then, to voluntarily producing the placebo effect lies in provoking, by whatever means, the psychological experience that corresponds to the physiological condition we wish to obtain.

In using the term correspondence, I do not mean to imply a dualistic relationship between the mind and body such that mental events are somehow *causing* physiological events. As stated in my introduction, I am merely making use of the basic identity thesis by which any mental state is assumed to have a corresponding bodily or brain state. This is essentially the same identity

[4]There need not be a one-to-one correspondence between the observed physiological effect and the patient's subjective experience, inasmuch as the target effect could arise anywhere along a chain of physiological events, the first of which having been triggered by a corresponding event in the mind. Dr. Ben Whatley brought the possibility of such upstream correspondence to my attention.

assumption in Kirsch (1985), Hyland (1985), Kirsch and Hyland (1987), and Hyland and Kirsch (1988) that served as the metatheoretical foundation for all of Kirsch's later work on response expectancies. In keeping with virtually all monist philosophies, Kirsch and Hyland assumed that for every mental state there is a corresponding brain state with which that mental state is associated. They also assumed that the "relation between a mental event and its physiological substrate is better described as an identity relation than as a relation of cause and effect" (Kirsch and Hyland, 1987, p. 421). Mental states do not *cause* physiological states, in other words, mental states *are* physiological states (and vice versa for the physiological correlates of mental states). A feeling of embarrassment does not cause the physiological activity with which it is associated; rather, the psychological experience of embarrassment and the physiological counterpart of this experience are two ways of describing the same event. We can speak of causal connections between mental states or causal connections between physiological states — based on the similar but independently conceived notions of causal isomorphism (Kirsch, 1985) and complementarity (Hyland, 1985) — but not of causal connections between these two categories of phenomena, at least not without invoking dualism and violating the law of conservation of energy (Kirsch, 1985). This view still allows for directionality between mental and physiological states, however. When alcohol is introduced into the nervous system, the cause of inebriation is clearly physiological; likewise, when someone chooses to have a drink, that choice can have any number of psychological causes behind it. The identity assumption adopted here presupposes that such brain/mind processes represent two sides of the same coin, regardless from which side they are initiated.

As I have already mentioned, Kirsch (1997) was not ready to extend this identity assumption to all placebo phenomena. To Kirsch et al. (2004), for example, it seems "highly unlikely that [Ader's] rats could expect immunosuppression or even have any representation of the phenomenon" (p. 385). This is a perfectly reasonable statement. Rats have no conception of the immune system, let alone the possibility of suppressing it with drugs, so it seems ridiculous to think they could have had any expectations regarding it. From the perspective of full correspondence, however, expectancy theory could still apply *if* the rats were led to expect, not the idea of a complex physiological phenomenon, but rather the feeling that corresponds to it. Unlike higher forms of consciousness, feelings are chiefly generated in the brain stem and thalamus rather than the more evolutionarily recent cerebral cortex; it is therefore reasonable to suppose that feelings are not restricted to humans or even to mammals (Damasio and Carvalho, 2013). Expectancy theory could thus broaden its range of application if it extended its investigations to feelings.

Of all our conscious experiences, feelings are the most likely correlates of type II physiological placebo effects. As defined by Damasio and Carvalho (2013),

"[f]eelings are mental experiences of body states" (p. 143). They include hunger, thirst, fear, and many varieties of pain and pleasure. As mental correlates of the body's physiological state they assist the organism in maintaining its internal homeostatic equilibrium. Because they are generated in the evolutionarily older regions of the brain, feelings are believed to represent the earliest forms of conscious experience (Damasio and Carvalho, 2013). If the bodily states of type II physiological placebo effects are also accompanied by feelings, it is possible that the ability to produce placebo effects is not a relatively recent evolutionary adaptation, as some have suggested (Bendesky and Sonabend, 2005; Humphrey, 2002; Trimmer, Marshall, Fromhage, McNamara, and Houston, 2013), but a rather ancient one. I first encountered this counterintuitive idea in Delboeuf's (1887) reflections on the origin of the curative powers of hypnosis, which are similar in many respects to those of placebos (Kirsch, 1999). Like Damasio and Carvalho (2013), Delboeuf believed that the conscious experiences of early organisms were restricted to the feelings associated with their internal bodily states. Over the course of evolution, Delboeuf went on to speculate, the regulation of these internal states became increasingly automated, allowing some organisms to concentrate their attention on the sensations produced by their developing sense organs and, eventually, on the thoughts generated by their evolving cognitive processes. Only, the ability to influence the processes that govern the internal bodily states was never lost, so that when the hypnotic subject heals himself, he is "reclaiming possession of a power he had ceased to exercise, but not abdicated" (Delboeuf, 1887, p. 812). On this view, many of the feelings associated with placebo effects would constitute a primordial record of our psychological past, comprising a wide range of psychological experiences, in sync with their corresponding physiological states, that have been pushed to the back of, but not expunged from, our modern cortex-dominated minds.

We may safely assume Ader's rats experienced the feeling of nausea; after all, it was for its nausea-inducing quality that Ader originally used cyclophosphamide. However, it is a different matter to assume that the rats experienced the psychological correlate of the immunosuppression produced by the drug. We do not know if they experienced it because we cannot ask them how they felt. But though we may not be able to test full correspondence on animals, we *can* test it on humans. And if the predicted consequences of full correspondence are not borne out by the empirical evidence in human subjects, then it is wrong. It does not matter whether one believes rats have psychological experiences or not, or even whether one thinks the theory of full correspondence is plausible or not. What matters is whether the novel results predicted by the theory are in fact observed or not.

Here is one such prediction. Full correspondence predicts similar subjective experiences when the same placebo-induced type II physiological effects are observed across patients. Suppose after receiving a placebo, 55% of the recipients

show a physically detectable placebo effect while the other 45% do not. The full correspondence model predicts that the 55% who responded to the placebo will report a psychological experience associated with the effect, while the 45% who did not respond to the placebo will report no such experience. Finding evidence along these lines in the literature is difficult, however, because researchers rarely report how their subjects feel when investigating type II physiological placebo effects.

And here we touch upon a profoundly important historical and philosophical issue. The reason placebo theorists rarely report how their subjects feel is because they rarely consider the possibility that introspective experience could be relevant to understanding type II physiological placebo effects. The reason, in turn, for this blind spot regarding conscious experience has been ongoing for centuries: over the course of our scientific training and professional careers, we have been led to internalize, in true Kuhnian fashion, the notion that the study of conscious experience is somehow not a legitimate scientific pursuit. In discussions with some of my colleagues, for instance, I have been told that full correspondence fails because there is something "unscientific" about it. They are correct; it is not, technically speaking, a scientific theory. But this is only true, not because of a limitation in the theory, but because of a limitation in our criteria for what counts as a scientific theory. According to the received view, consciousness is a phenomenon to be explained away, rather than a source of evidence for explaining phenomena. Consciousness is supposed to be the *explanandum*, not part of the *explanans*. But if the full correspondence interpretation is correct, the placebo effect will remain impossible to understand so long as consciousness is not part of the explanation.

In *The Feeling of What Happens*, from which I borrowed the opening title for this paper, the renowned neurologist and clinician, Antonio Damasio, wrote how "[s]tudying consciousness was simply not the thing to do before you made tenure, and even after you did it was looked upon with suspicion. Only in recent years has consciousness become a somewhat safer topic of scientific inquiry" (1999, p. 7).[5] Some ten years later, it has fortunately become a somewhat safer topic for placebo theorists as well. Kaptchuk et al. (2009), for example, recently carried out a qualitative investigation of the subjective experiences of

[5]The turning point in the legitimization of the study of consciousness is marked by two important publications, both in 1994: *The Astonishing Hypothesis: The Scientific Search for the Soul,* by Francis Crick, Nobel laureate and co-discoverer of the DNA molecule; and the first volume of the *Journal of Consciousness Studies*. The first issue of the journal began with an interview with Crick and included articles by several eminent scholars interested in the study of consciousness. Although the topics ranged from the binding problem and quantum theories of mind to machine consciousness and mystical experiences, the articles shared the same underlying assumption: conventional approaches having failed to solve deep long-standing problems in consciousness, the time had come to take consciousness more seriously and to propose methods of inquiry better suited to understanding it.

patients undergoing placebo acupuncture for irritable bowel syndrome. As far as this research team knew, this is the first time anyone had analyzed the experiences of placebo patients in a randomized control trial. Indeed, given the existence of multiple competing theories of the placebo effect, they noted how peculiar it was that "none has been informed by actual interviews of patients undergoing placebo treatment" (p. 382). As a further sign of the times, another recent placebo study of irritable bowel syndrome patients pointed to the same lacuna, adding that to the authors' knowledge, theirs was "the first study to directly compare patients' experience of a placebo treatment versus an active treatment" (Vase, Nørskov, Petersen, and Price, 2011, p. 1917).

The results of the latter study, which included administering rectal placebo during a painful rectal balloon distention procedure, were consistent with those predicted by full correspondence: they showed placebo responders "actively engaging in generating a mindset for pain reduction," which, once established during the first 20 minutes following administration of placebo, maintained itself with less deliberate mental effort during the next 20 minutes (Vase et al., 2011, p. 1919). It seems once the placebo recipient settles into a state of pain reduction, it becomes easier, almost effortless for some, to prolong that state of mind. This is one example, incidentally, of the kind of fruitful research results one would expect to find under a full correspondence paradigm. The subjective measures of pain reduction in this study are corroborated, moreover, by objective measures of pain reduction in a previous fMRI study of irritable bowel syndrome using a similar design. Price, Craggs, Verne, Perlstein, and Robinson (2007) found reduced activity in the pain-related areas of the brain in irritable bowel syndrome patients who received rectal placebo during the rectal distention procedure. A stronger test of full correspondence would of course combine the above two studies into one, so that the subjective experience of placebo recipients could be directly compared by fMRI.

Ideally, what we need to assess in the theory of full correspondence are studies that compare the subjective experiences of patients who respond to type II placebos in a physiologically measurable way with those who do not. It so happens this is precisely the kind of study undertaken by Benedetti et al. (2004) in a surgical experiment involving Parkinson patients. The object of the experiment was not to test the theory of full correspondence, of course, but rather to see whether placebo medication for Parkinson patients could influence the activity of the brain and produce clinical improvement. But, as we shall see, Benedetti et al. included a condition that makes it possible to test full correspondence: they asked the placebo recipients how they felt.

A group of 11 Parkinson patients were administered three injections of apomorphine, a potent antiparkinsonian drug, in the days leading up to surgery. The surgical procedure consisted of inserting electrodes into the subthalamic nucleus, a region of the brain important in the treatment of Parkinson's disease,

and recording the neuronal activity before and after administration of placebo apomorphine. When the patients received their placebo injection, they were told it was the same apomorphine as the days before and that a feeling of well-being would follow. The effect of the placebo injection was measured in three ways: (1) degree of arm rigidity, (2) level of subthalamic nucleus neuronal activity, and (3) type of subjective experience. This experiment is unusual in that physiological measures are rarely paired with subjective measures when investigating placebo effects, especially type II physiological placebo effects. Indeed, one of the reasons Benedetti et al. (2004) included this subjective measure was to challenge the frequent objection that when patients report feeling better after placebo administration, such reports correspond to the patient's biases, "such as the patient's desire to please the investigator," rather than to objective physiological changes (Benedetti, 2009, p. 38). It is worth noting that the experimenters took great care not to influence their patients' introspective reports, so that the neurologist who recorded them had no knowledge of the patients' performance on the muscular and neuronal evaluations. The experimenters found that the six placebo recipients who displayed the physiological effects of apomorphine, namely decreased arm rigidity and reduced subthalamic nucleus activity, *were the same six to report feelings of well-being*, whereas the five non-responders neither displayed these physiological effects nor reported experiencing them. For example, the placebo responders reported such things as, "I'm falling asleep, like after apomorphine," "I feel like after the usual therapy," or "I feel much better," whereas the non-responders' reports were completely negative, such as, "I don't feel any effect," "It doesn't work," or "I feel no change" (p. 587). In other words, just as full correspondence leads us to expect, not only did the placebo responders reproduce the physiological effects of the drug, they also shared subjective experiences that roughly corresponded to those effects, while the non-responders neither reproduced nor felt these effects. Of course, full correspondence does not replace the conditioning model of the placebo effect; rather, it integrates the conditioning model with other models of the placebo effect by positing a common feature, namely, the patient's subjective experience. Also, with respect to my suggestion that feelings are the most likely subjective correlates of type II physiological placebo effects, it is interesting to note that the main targets of apomorphine in this experiment were the striatum and the subthalamic nucleus, both of which are common to all vertebrates and, therefore, extremely old from the point of view of evolution.

Case Study: Conditioning with or without Conscious Cognition

Let us apply full correspondence to a placebo study by Benedetti et al. (1999) that is widely believed to have occurred in the absence of conscious cognition (Benedetti, 2009; Kirsch, 2004; Moerman, 2002b; Stewart–Williams and Podd,

2004). In this conditioning study, surgical patients were given open injections of buprenorphine on the two days following their operation and a placebo injection on the third; thus, a conditioning paradigm very similar to the one Benedetti et al. (2004) would employ five years later, except this time the medication was administered after rather than before surgery.

Buprenorphine is a powerful semi-synthetic narcotic that can depress respiratory volume by 15% to 20% when taken in clinical doses (in this case 0.2 mg). It is important to note that respiratory depression is a typical side effect of narcotics and usually goes unnoticed by patients. When the placebo was administered, the patients were told it was the same drug they had received on the previous two days. As predicted, the experimenters found that both the buprenorphine and the placebo produced a significant drop in respiration. "Interestingly," Benedetti (2009) later wrote, "the patients themselves did not expect any effect and did not notice any decrease in ventilation, which suggests this effect is an *unconscious conditioning mechanism* [emphasis added] whereby the act of giving the drug was the conditioned stimulus" (p. 184). In addition to expectancy theory, these results seem to rule out meaning theory too, as Moerman (2002b) wrote regarding the experiment: "The treatments clearly had meaning ('narcotics are powerful painkillers'), but they did not have the meaning 'narcotics repress respiration,' even though that's true" (p. 124).

Full correspondence invites a different interpretation, one in which expectancy theory and the meaning response are not so easily dismissed. Supposing the theory of full correspondence is true, then the placebo reproduced not only the physiological effects of buprenorphine in placebo responders, but the psychological effects as well. This is hardly a controversial assumption given that Benedetti et al. (2004) found a firm match between the psychological effects of apomorphine and their corresponding physiological effects in the Parkinson study described above. It is therefore possible that the placebo injection led patients to expect they would feel the sensations associated with buprenorphine, which had the *meaning*: "buprenorphine is a powerful painkiller *and* produces a peculiar feeling," which in turn provoked the physiological and psychological effects of buprenorphine, and thereby the side effect of respiratory depression. In other words, that the placebo responders did not notice the effect the injections had on their respiration does not rule out the possibility that this placebo-induced side effect was mediated by expectation or the meaning response. After all, should we really be surprised, if, after manipulating expectations, a placebo aspirin produced an anti-inflammatory response in someone who knows nothing of its anti-inflammatory properties? If the subjective effects of aspirin are reproduced, full correspondence predicts that the physiological effects will be reproduced as well, including reduced inflammation. Under a full correspondence paradigm, expectation and meaning are therefore still theoretically possible in the Benedetti et al. (1999) study because what patients are expect-

ing and what buprenorphine means to them is that they will *feel* a certain way, and that feeling comes with its own physiological concomitants regardless of whether the patients are aware of them or not. But even if expectation and meaning played no part in the study, this would not imply that the placebo-induced physiological effects had no corresponding effects in the mind. Again, based on the results of Benedetti et al. (2004), it would be surprising if the physiological effects of buprenorphine were not accompanied by their psychological counterparts (they accompanied apomorphine, why not buprenorphine). Full correspondence does not guarantee expectation or meaning played a part in Benedetti et al. (1999), but it does suggest that theorists need not assume, as every one has, that this placebo experiment occurred in the absence of a mental experience, cognitive or otherwise.

Conclusion: A Meta-theoretical Framework

Full correspondence provides the conceptual means with which to resolve the conflict between meaning-oriented and mechanism-oriented approaches. The main source of the conflict is that certain placebo phenomena, such as conditioned immunosuppression in rats, are apparently so completely governed by mechanical processes that consciousness seems absent from the causal core of the placebo effect. And if consciousness is absent, so are expectation, meaning, and culture. I have argued, however, that placebo effects could occur through conditioning and yet also be *felt*, as was the case with the effects of placebo apomorphine in patients suffering from Parkinson's disease (Benedetti et al., 2004). Hence, if the type II physiological placebo effects produced by other conditioning experiments are similarly accompanied by corresponding mental experiences, then the epistemological gap between meaning and mechanism is effectively closed. According to the theory of full correspondence, in other words, the common denominator on both sides of the meaning–mechanism divide is the mental experience that accompanies all placebo effects.

Conditioning theory is correct to point out that many placebo effects can be explained through conditioned learning, but we should not necessarily assume that some conditioning procedures produce placebo effects in the absence of a felt experience. As pointed out in my discussion of Benedetti et al. (1999), the placebo responders could have felt the physiological effects of placebo buprenorphine even if these effects were not mediated by expectation, a meaning response, or some other higher cognitive process. In other cases, meaning theory and other anthropological approaches are correct in emphasizing the cultural factors that influence the placebo effect, but they could strengthen their case by attending to how the placebo effect is subjectively experienced. By interviewing placebo responders and determining the content of their mental life (or at least suitable proxies of that mental life, since interviews and questionnaires

can only provide analogues of first person experiences, not the experiences per se), as Kaptchuk et al. (2009) have done in their pioneering study, social theorists could establish more precisely how culture and meaning shape and give rise to certain placebo effects. Expectancy theory could similarly increase its explanatory power by exploiting the implications of full correspondence. Unlike culturally based theories, expectancy theory concentrates its attention not on the social causes of the placebo effect, but on one of its psychological causes. It is therefore closer to the source of the action, but it stops short of the placebo's final denouement. Expectancy theory has hitherto focused on the state of the subject's mind *before* the placebo effect occurs, rather than *while* it is occurring. It has been chiefly concerned with the final steps leading up to the effect, not the effect itself.

The theory of full correspondence neither replaces nor competes with the various existing approaches to the placebo effect; it is a meta-theory, designed to unify mechanically-oriented approaches and meaning-oriented approaches within the same epistemological and ontological framework. Nor does it establish the superiority of one approach over another. As there is not one but several ways of producing placebo effects, there are also several ways of describing how they are produced. In some cases expectation is the dominant cause, in others it is conditioning, and in still others it is meaning, hope, belief, or a combination thereof. Like so many specialized engineers, each approach is best suited for its particular area of expertise, its particular way of tunneling into Hering's metaphorical mountain. But they all meet at the same point: the place where the meaning and the mechanics of the placebo effect coincide with the feeling of what happens.

References

Ader, R. (1974). [Letter to the editor]: Behaviorally conditioned immunosuppression. *Psychosomatic Medicine, 36,* 183–184.

Ader, R. (1997). The role of conditioning in pharmacotherapy. In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 138–165). Cambridge, Massachusetts: Harvard University Press.

Ader, R., and Cohen. N. (1975). Behaviorally conditioned immunosuppression. *Psychosomatic Medicine, 37,* 333–340.

Barrett, B., Muller D., Rakel D., Rabago D., Marchand, L., and Scheder, J.C. (2006). Placebo, meaning and health. *Perspectives in Biology and Medicine, 49,* 178–198.

Bendesky, A., and Sonabend, A.M. (2005). On Schleppfuss' path: The placebo response in human evolution. *Medical Hypotheses, 64*(2), 414–416.

Benedetti, F. (2009). *Placebo effects: Understanding the mechanisms in health and disease*. New York: Oxford University Press.

Benedetti, F., Amanzio, M., Baldi, S., Casadio, C., and Maggi, G. (1999). Inducing placebo respiratory depressant responses in humans via opioid receptors. *European Journal of Neuroscience, 11,* 625–631.

Benedetti, F., Colloca, L., Torre, E., Lanotte, M., Melcarne, A., Pesare, M., et al. (2004). Placebo-responsive Parkinson patients show decreased activity in single neurons of subthalamic nucleus. *Nature Neuroscience, 7,* 587–588.

Branthwaite, A., and Cooper, P. (1981). Analgesic effects of branding in treatment of headaches. *British Medical Journal, 282,* 1576–1578.
Brody, H. (1997). The doctor as therapeutic agent: A placebo effect research agenda. In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 77–92). Cambridge, Massachusetts: Harvard University Press.
Crick, F. (1994). *The astonishing hypothesis: The scientific search for the soul.* New York: Maxwell Macmillan International.
Damasio, A. (1999). *The feeling of what happens. Body and emotion in the making of consciousness.* San Diego: Harvest Book.
Damasio, A. (2002). How the brain creates the mind. *Scientific American Special Edition, 12*(1), 4–9.
Damasio, A., and Carvalho, G.B. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews Neuroscience, 14,* 143–152.
de Craen, A.J.M., Roos P.J., Leonard de Vries, A., and Kleijen, J. (1996). Effect of colour of drugs: Systematic review of perceived effect of drugs and of their effectiveness. *British Medical Journal, 313*(7072), 1624–1626.
de Craen, A.J.M., Tijssen, J.G.P., de Gens J., and Kleijnen, J. (2000). Placebo effect in the acute treatment of migraine: Subcutaneous placebos are better than oral placebos. *Journal of Neurology, 247,* 183–188.
Delboeuf, J. (1887). De l'origine des effets curatifs de l'hypnotisme. *Bulletins de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique,* 57th year, 3rd series, *13*, 773–812.
Delboeuf, J. (1993). Louise Lateau. In J. Carroy and F. Duyckaerts (Eds.), *Le Sommeil et les rêves, et autres textes* (pp. 387–400). Paris: Fayard. (Original work published in 1869)
Di Blasi, Z., Harkness, E., Ernst, E., Georgiou, A., and Kleijnen, J. (2001). Influence of context effects on health outcomes: A systematic review. *Lancet, 357*(9258), 757–762.
Flaten, M.A., Aasli, O., and Blumenthal, T.D. (2003). Expectations and placebo responses to caffeine-associated stimuli. *Psychopharmacology, 169,* 198–204.
Hahn, R.A. (1985). A sociocultural model of illness and healing. In L. White, B. Tursky, and G. E. Schwartz (Eds.), *Placebo: Clinical phenomena and new insights* (pp. 167–195). New York: Guilford Press.
Hahn, R.A. (1995). *Sickness and healing: An anthropological perspective*. New Haven: Yale University Press.
Harrington, A. (1997). Introduction. In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 1–11). Cambridge, Massachusetts: Harvard University Press.
Hróbjartsson, A., and Gøtzsche, P.C. (2001). Is the placebo powerless? — An analysis of clinical trials comparing placebo with no treatment. *The New England Journal of Medicine, 344,* 1594–1602.
Hull, J.G., and Bond, C.F. (1986). Social and behavioral consequences of alcohol consumption and expectancy: A meta-analysis. *Psychology Bulletin, 99,* 347–360.
Humphrey, N. (2002). *The mind made flesh: Essays from the frontiers of psychology and evolution.* Oxford: Oxford University Press.
Hussain, M.Z., and Ahad, A. (1970). Tablet colour in anxiety states. *British Medical Journal, 3*(5720), 466.
Hyland, M.E. (1985). Do person variables exist in different ways? *American Psychologist, 40,* 1003–1010.
Hyland, M.E., and Kirsch, I. (1988). Methodological complementarity: With and without reductionism. *Journal of Mind and Behavior, 9,* 5–12.
Jensen, M.P., and Karoly, P. (1991). Motivation and expectancy factors in symptom perception: A laboratory study of the placebo effect. *Psychosomatic Medicine, 53,* 144–152.
Jensen, K.B., Kaptchuk, T.J., Kirsch, I., Raicek, J., Lindstrom, K.M., Berna, C., et al. (2012). Nonconscious activation of placebo and nocebo pain responses. *Proceedings of the National Academy of Sciences, 109*(39), 15959–15964.
Kaptchuk, T.J., Shaw, J., Kerr, C.E., Conboy, L.A., Kelley, J.M., Csordas, T. J., et al. (2009). "Maybe I made up the whole thing": Placebos and patients' experiences in a randomized controlled trial. *Culture, Medicine and Psychiatry, 33,* 382–411.
Kienle, G.S., and Kiene, H. (1997). The powerful placebo effect: Fact or fiction? *Journal of Clinical Epidemiology, 50*(12), 1311–1318.

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist, 40,* 1189–1202.
Kirsch, I. (1997). Specifying nonspecifics: Psychological mechanisms of placebo effects. In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 166–186). Cambridge, Massachusetts: Harvard University Press.
Kirsch, I. (1999). Hypnosis and placebos: Response expectancy as a mediator of suggestion effects. *Anales de psicología, 15*(1), 99–110.
Kirsch, I. (2004). Conditioning, expectancy, and the placebo effect: Comment on Stewart–Williams and Podd. *Psychology Bulletin, 130,* 341–343.
Kirsch, I., and Hyland, M.E. (1987). How thoughts affect the body: A metatheoretical framework. *Journal of Mind and Behavior, 8,* 417–434.
Kirsch, I., Lynn, S.J., and Miller, R. (2004). The role of cognition in classical and operant conditioning. *Journal of Clinical Psychology, 60*(4), 369–392.
Kirsch, I., and Sapirstein, G. (1999). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medications. In I. Kirsch (Ed.), *How expectancies shape experience* (pp. 303–320). Washington, DC: APA.
Kirsch, I., and Weixel, L.J. (1988). Double-blind versus deceptive administration of a placebo. *Behavioral Neuroscience, 102*(2), 319–323.
Kleinman, A. (1986). *Social origins of distress and disease: Depression, neurasthenia, and pain in modern China.* New Haven: Yale University Press.
Kleinman, A. (1998). "Sociomatics": The contributions of anthropology to psychosomatic medicine. *Psychosomatic Medicine, 60,* 389–393.
Kradin, R. (2004). The placebo response: Its putative role as a functional salutogenic mechanism of the central nervous system. *Perspectives in Biology and Medicine, 47*(3), 328–338.
Moerman, D.E. (2002a). Explanatory mechanisms for placebo effects: Cultural influences and the meaning response. In H.A. Guess, A. Kleinman, J.W. Kusek, and L.W. Engel (Eds.), *The science of the placebo: Toward an interdisciplinary research agenda* (pp. 35–52). London: British Medical Journal Books.
Moerman, D.E. (2002b). *Meaning, medicine and the "placebo effect."* Cambridge: Cambridge University Press.
Montgomery, G., and Kirsch, I. (1996). Mechanisms of placebo pain reduction: An empirical investigation. *Psychological Science 7*(3), 174–176.
Montgomery, G., and Kirsch, I. (1997). Classical conditioning and the placebo effect. *Pain, 72,* 107–113.
Phillips, D.P., Ruth, T.E., and Wagner, L.M. (1993). Psychology and survival. *Lancet, 342*(8880), 1142–1145.
Price, D.D., Craggs, J., Verne, G.N., Perlstein, W.M., and Robinson, M.E., (2007). Placebo analgesia is accompanied by large reductions in pain-related brain activity in irritable bowel syndrome patients. *Pain, 127,* 63–72.
Rickels K., Hesbacher, P.T., Weise, C.C., Gray, B., and Feldman, H.S. (1970). Pill and improvement: A study of placebo response in psychoneurotic outpatients. *Psychopharmacologia, 16*(4), 318–328.
Stewart–Williams, S., and Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin, 130,* 324–340.
Trimmer, P.C., Marshall, J.A.R., Fromhage, L., McNamara, J.M., and Houston, A.I. (2013). Understanding the placebo effect from an evolutionary perspective. *Evolution and Human Behavior, 34*(1), 8–15.
Turner, S.R. (1994). *In the mind's eye: Vision and the Helmholtz–Hering controversy*. Princeton: University of Princeton Press.
Vase, L., Nørskov, K. N., Petersen, G.L., and Price, D.D. (2011). Patients' direct experiences as central elements of placebo analgesia. *Philosophical Transactions of the Royal Society B, 366,* 1913–1921.
Waber, R.L., Shiv, B., Carmon, Z., and Ariely, D. (2008). Commercial features of placebo and therapeutic efficacy. *Journal of the American Medical Association, 299*(9), 1016–1017.

Critical Notices
Book Reviews
Book Notes

The Journal of Mind and Behavior
Summer 2014, Volume 35, Number 3
Pages 185–188
ISSN 0271–0137

The Peripheral Mind: Philosophy of Mind and the Peripheral Nervous System. István Aranyosi. Oxford, United Kingdom: Oxford University Press, 2013, 256 pages, $60.00 hardcover.

Reviewed by Michael Madary, Universität Mainz

Much of the action and excitement in the philosophy of mind over the last couple of decades has been in a movement to look beyond the brain for locating and explaining mental states. This movement consists in a number of different claims. We have heard, for instance, that the mind extends into artifacts, and that the mind is brought forth or enacted or constituted by the active living body. In his recent book, *The Peripheral Mind*, István Aranyosi defends a neglected middle ground in the debate, a middle ground between the brain and the external world. Aranyosi urges that we take seriously the peripheral nervous system in our investigation into the mind. More specifically, the main thesis of his book is the peripheral mind hypothesis, which is that "Conscious mental states typically involved in sensory processes are partly constituted by subprocesses occurring at the level of the [peripheral nervous system]" (p. 22).

I find the book overall to be thought-provoking, especially as it brings a fresh perspective on a number of issues in contemporary philosophy of mind, including semantic externalism and some issues in neuroethics. One attraction of the book is Aranyosi's ecumenical methodology; he draws from cultural anthropology, detailed neurophysiology, illusions of embodiment, continental phenomenology, thought experiments (Stinky Earth is my favorite of these), and even his own personal experiences, which are directly relevant. Due to the scope of the book, I must leave out quite a bit in my discussion. My focus will be on its main thesis, which is original and potentially relevant in a wide range of issues, as Aranyosi indicates. The central argument for the main thesis can be found in the seventh chapter of the book. As I explain below, I find the argument lacking.

Before looking at the argument for the peripheral mind hypothesis, I should locate the claim within the existing literature. Probably the most important objection to the various theses advocating extra-cranial extension of the mind is the objection that its proponents fail to appreciate the distinction between causation and constitution (Adams and Aizawa, 2008; Block, 2005; Prinz, 2006). The objection is that proponents of extension identify important causal contributions to mental states and then fallaciously

---

Correspondence concerning this review should be addressed to Dr. Michael Madary, Johannes Gutenberg–Universität Mainz, FB05 Philosophie und Philologie, Jakob Welder Weg 18, 55099 Mainz, Germany. Email: madary@uni-mainz.de

conclude that these causal contributions actually constitute those mental states. Aranyosi is aware of the distinction, and the objection. Given this state of affairs, it is crucial for his defense of the peripheral mind hypothesis to make a clear case for the constitutive claim, a case why the peripheral nervous system makes a constitutive, rather than a "merely" causal, contribution to conscious mental states.

The main basis for the constitutive claim is a number of empirical results having to do with illusions of embodiment. Aranyosi begins with Aristotle's illusion: cross the index and middle fingers, then touch the tips of both crossed fingers simultaneously with a pencil. (Hold the pencil perpendicular to your crossed fingers, and place the pencil in the "V" created by your crossed fingertips.) Many people experience being touched by two objects, despite the visual percept (and veridical belief) that they are being touched by one object. Aranyosi then moves on to describe a number of other illusions involving proprioception and touch, including a variation on the rubber hand illusion and his own variation on Aristotle's illusion. One key experimental finding for Aranyosi's argument is that the tactile illusions can be lost for subjects who spend a long time with crossed fingers (Benedetti, 1991). He reaches the plausible conclusion that the tactile properties of our fingertips depend on the history of the ways in which they have been stimulated by objects (p. 134).

With these empirical results in place, Aranyosi goes on to apply a counterfactual causal analysis of the illusory experience in order to justify the constitutive claim. He suggests that the "one causal contributor" to the illusory experience is the absence of a particular kind of stimulation history (p. 135). Counterfactually: if the stimulation history had been different, there would have been no illusory experience. Aranyosi concludes that, since the actual stimulation of the fingers "is a contributor to my paradoxical experience," then ". . . we should understand these [peripheral nervous system] processes as constitutive contributors to the experience" (p. 135).

Now I will offer some critical remarks, beginning with the argument just sketched. The main problem that I find with this argument is that it depends on a dubious background assumption. The implicit assumption is that a counterfactual analysis reveals the "only one causal contributor" to an event (ibid.). This assumption is questionable because it is plausible that many events have multiple causes that can be revealed using a counterfactual analysis. In this case, I suggest, the actual stimulation of the peripheral nerves is a good candidate for another causal contribution to the experience. It is wrong to suppose, as Aranyosi seems to do, that all events have one single cause and that all other contributing factors are constitutive. Instead, one could plausibly maintain that the other contributing factors are background causes. Another relevant point here is that counterfactual causal analyses have been used as ways to model commonsense judgments about causation. In this case, the counterfactual analysis is used to reach a decidedly non-commonsense judgment about the cause of an event. Thus Aranyosi's argument may raise a problem for counterfactual analyses of causation rather than support a conclusion about the peripheral nervous system.

Part of the difficulty here might lie in the fact that the causal/constitutive distinction is a poor fit for theorizing in empirical science. Following Ross and Ladyman (2010), the root problem in the debate is that the causal/constitutive distinction belongs to analytic metaphysics (or, less charitably, to folk physics), but it is being applied to a theoretical dispute in the empirical sciences of the mind. According to Ross and Ladyman, since the distinction has no place in the mature sciences such as physics and chemistry, it should find no place in the sciences of the mind. Instead of making the constitutive claim, then, one could instead argue that our best scientific models of the mind are

those that include, in this case, the peripheral nervous system. I suspect that elements of Aranyosi's book could be adapted to this claim, though I will not pursue the issue.

Apart from the relevance of constitutive claims for the sciences of the mind, I'd like to raise two further worries about the peripheral mind hypothesis. The first worry is that Aranyosi excludes dreams from his hypothesis, because "the connection between sensory states in dreams and the [peripheral nervous system] is much less tight in actual fact" (p. 22). Since dreams have already been raised in the debate over whether the conscious mind is partly constituted by extra-cranial processes (Block, 2005; Noë, 2004: chapter 7), I was somewhat surprised to see their casual dismissal here. More to the point, if we can have a phenomenal state in a dream, without the constitutive (or even causal) role of the peripheral nervous system, then we have strong prima facie reasons for thinking that the peripheral nervous system is not constitutive of particular phenomenal states. It would seem that such a conclusion would be in tension with the peripheral mind hypothesis.

A second worry is that Aranyosi does not address evidence for the plasticity of the body schema. There is experimental evidence that tool use can change the receptive field properties of the cortical neurons that play a role in body representation (see Maravita and Iriki, 2004 for a review of the literature). This evidence suggests that our body representations are mostly determined by the central nervous system, and that the peripheral nervous system may not play a significant role. For instance, assume that my body representation can become extended when I am using a rake, such that the tip of the rake is represented as the tip of my limb. Also assume, in accordance with the experimental findings, that this extension is due to the plasticity of neuronal activity in the central nervous system. In such a case, it is not clear to me that the properties of the peripheral nervous system are of any explanatory interest — it is not as if the peripheral nervous system itself extends into the rake. Perhaps the peripheral nervous system will be important for an explanation of the plasticity of body representation, but the onus is on proponents of the peripheral mind hypothesis to make that case.

Overall, *The Peripheral Mind* has the virtues of originality and scope. But the trade-off for scope is slow and careful argumentation, as I indicated using the example of the main argument of the book.

References

Adams, F., and Aizawa, K. (2008). *The bounds of cognition.* West Sussex, United Kingdom: Blackwell Publishing.

Benedetti, F. (1991). Perceptual learning following a long lasting tactile reversal. *Journal of Experimental Psychology, 17,* 267–277.

Block, N. (2005). [Review of Alva Noë's] *Action in Perception*. *Journal of Philosophy, 5,* 259–272.

Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences, 2,* 79–86.

Noë, A. (2004). *Action in perception*. Cambridge, Massachusetts: MIT Press.

Prinz, J. (2006). Putting the breaks on enactive perception. *Psyche, 12,* 1–19.

Ross, D., and Ladyman, J. (2010). The alleged coupling-constitution fallacy and the mature sciences. In R. Menary (Ed.), *The extended mind* (pp. 155–166). Cambridge, Massachusetts: MIT Press.

A NOTE ON OUR BOOK REVIEW POLICY

We will accept book reviews for publication each issue. Authors wishing to submit book reviews are urged to write with the above interdisciplinary framework firmly in mind. All books *solicited* from publishers will be sent to selected individuals for review. JMB also accepts unsolicited reviews. Reviews should be absent of all titles except the name of the work reviewed, author of work reviewed, place of publication, publisher, date of latest publication, number of pages, and cost. Any individual wishing to submit a review should contact our Book Review Editor for further information: Steven E. Connelly, Ph.D., Department of English, Indiana State University, Terre Haute, Indiana 47809. Email: sconnelly@isugw.indstate.edu

JMB is abstracted or indexed in: *Cultures, Langues, Textes: La Revue de Sommaires; Current Contents (Social and Behavioral Sciences); EMBASE/Excerpta Medica; International Bibliography of Book Reviews; International Bibliography of Periodical Literature; Linguistics and Language Behavior Abstracts; Physics Abstracts; Psychiatric Rehabilitation Journal; PsychINFO/Psychological Abstracts; Research Alert; Social Science Citation Index; Social Work Abstracts; Sociological Abstracts; The Philosopher's Index.* The Journal of Mind and Behavior website is located at www.umaine.edu/jmb/.

The Journal of Mind and Behavior
Summer 2014 Vol. 35 No. 3

CONTENTS