
Consciousness as Intransitive Self-Consciousness: Two Views and an Argument

URIAH KRIEDEL
Brown University
Providence, RI 02912
USA

I Introduction: Consciousness and Intransitive Self-Consciousness

The word 'consciousness' is notoriously ambiguous. This is mainly because it is *not* a term of art, but a mundane word we all use quite frequently, for different purposes and in different everyday contexts. In this paper, I am going to discuss consciousness in one specific sense of the word. To avoid the ambiguities of the word 'consciousness,' I will introduce a term of art: *intransitive self-consciousness*. As the term suggests, the phenomenon I have in mind is a kind of self-consciousness, or self-awareness.¹

Intransitive self-consciousness is to be distinguished from *transitive* self-consciousness. One way to frame the distinction is brought out in the following pair of reports:

- (a) *x* is self-conscious of her thought that *p*.
- (b) *x* is self-consciously thinking that *p*.

1 In this paper, I use 'self-awareness' and 'self-consciousness' interchangeably.

The two statements do not report the same mental state. (a) reports the occurrence of a second-order state of self-consciousness whose *object* is *x*'s thought that *p*; (b) reports the occurrence of a first-order thought whose object is the state of affairs *that p*, and ascribes to it the property of occurring self-consciously. I call the mode of self-consciousness referred to in (a) 'transitive self-consciousness' and that referred to in (b) 'intransitive self-consciousness'.²

What is involved in a mental state being intransitively self-conscious? That is, what is it for a subject to have a mental state self-consciously? An acute description is offered by Alvin Goldman:

[Consider] the case of thinking about *x* or attending to *x*. In the process of thinking about *x* there is already an implicit awareness that one is thinking about *x*. There is no need for reflection here, for taking a step back from thinking about *x* in order to examine it... When we are thinking about *x*, the mind is focused on *x*, not on our *thinking* of *x*. Nevertheless, the process of thinking about *x* carries with it a non-reflective self-awareness.³

When we have a mental state self-consciously, there is a subtle awareness of self implicit in that state, whereby we are aware of ourselves as its *owners*. This is not just a matter of the traditional issue of ownership; it is a matter of ownership-awareness. It is not just that we *are* the owners; we are *aware* of being the owners, albeit implicitly and inattentively. To say that *x* has mental state M self-consciously, then, is to say that *x* is implicitly aware of her *having* M, or of M being *hers*.⁴

Suppose, for instance, that you suddenly hear a distant bagpipe. In your auditory experience of the bagpipe you are aware primarily, or *explicitly*, of the bagpipe sound; but you are also *implicitly* aware that this auditory experience of the bagpipe is *your* experience. That is, you are aware of yourself as the subject of experience. Similarly, when you are thinking that the almond trees are finally blooming again, you are also implicitly aware that this is what you are thinking. You do not need to explicitly reflect on your experience or thought in order to be thus aware

2 I call the phenomenon denoted in (a) transitive self-consciousness because 'to be self-conscious of' is a transitive verb. 'Self-consciously' is not a verb at all, but an adverb, but I call the phenomenon it is used to denote intransitive self-consciousness mainly to mark it off from transitive self-consciousness.

3 A. Goldman, *A Theory of Action* (Princeton, NJ: Princeton University Press 1970), 96

4 It is possible, of course, for *x* to have M without being aware that she is having M, or without being aware of M as hers, but then *x* is not having M *self-consciously*. So in the sense of intransitive self-consciousness, M would be non-conscious.

of them. That is, you do not need to enter a state of transitive self-consciousness.

There are many psychological and phenomenological differences between transitive and intransitive self-consciousness. Let me mention only four of them. First, a transitively self-conscious state is *introspective*, in that its object is always one of the subject's own mental states, whereas an intransitively self-conscious state is ordinarily not introspective, in that usually its object is an external state of affairs. Second, transitive self-consciousness is *rare*, in that our stream of consciousness takes explicit notice of itself relatively infrequently, whereas intransitive self-consciousness is *ubiquitous*, in that at any single moment of our waking life, we have at least one intransitively self-conscious state, so there is a dim self-awareness constantly humming in the background of our stream of consciousness. Third, entering a transitively self-conscious state is ordinarily a *voluntary* matter — we can choose to introspect — whereas the intransitive self-consciousness built into our experiences is not voluntary (we do not choose to have it); indeed, it is *involuntary* (we cannot 'shut it down' at will). Fourth, transitive self-consciousness is for the most part *effortful*, in that one must concentrate on one's inner thoughts and feelings, whereas intransitive self-consciousness is *effortless*, in that despite accompanying our conscious life permanently, it makes negligible demands on our cognitive resources. This is similar to the difference between, say, calculating and seeing: the former requires an effort on one's part, whereas the latter is effortless — it just happens to one.

(It is important to note, however, that states of transitive self-consciousness are often intransitively self-conscious as well. Thus, when I try to remember how I felt when I had my first kiss, I am *explicitly* aware of how I felt, but I am also *implicitly* aware of myself engaged in remembering how I felt. Since intransitive self-consciousness is ubiquitous, and accompanies us throughout our waking life, it is present also in states of transitive self-consciousness.)

My contention is that intransitive self-consciousness captures one of the mundane senses of the word 'consciousness.' We often use the adjective 'conscious' to indicate the presence of some sort of minimal self-awareness. This paper will discuss two competing accounts of consciousness as intransitive self-consciousness.

Some readers may feel that intransitive self-consciousness does not capture any important sense of 'consciousness.' It is possible, in a liberal mindset, to treat intransitive self-consciousness as a form of consciousness, but this form is unrelated to the 'real' sense of 'consciousness,' the sense of phenomenal consciousness — the sense in which there is *something it is like* for the subject to be in a conscious state (to use Nagel's oft-worn phrase).

In my opinion, this would be a misguided feeling: intransitive self-consciousness is absolutely central to our folk-psychological concept of consciousness. To my ear, there is something artificial in calling a mental state conscious when the subject is wholly unaware of its occurrence. Conscious states are not sub-personal states, which we may have, as it were, unawares. Furthermore, it is unlikely there could be anything it is like for a subject to be in a mental state she is unaware of being in. It would seem, then, that intransitive self-consciousness is a *necessary condition* for phenomenal consciousness: unless M is intransitively self-conscious, there is nothing it is like to be in M, and therefore M is not a phenomenally conscious state.⁵

Some readers may have a more radical feeling, that the notion of intransitive self-consciousness is empty, i.e., that there is no such thing as a non-reflective mode of self-awareness.⁶ To my mind, this feeling is tempting, partly, precisely because intransitive self-consciousness is so effortless yet ubiquitous. But unfortunately, I cannot here argue for the existence of intransitive self-consciousness. Let me only point out that the existence of some primordial form of self-awareness has been all but taken for granted throughout modern philosophy. Thus, Locke (*Essay Concerning Human Understanding*, II.xvii.19) states that it is 'impossible

5 To be sure, much more can — and should — be said about the relationship between intransitive self-consciousness and phenomenal consciousness. In particular, my claim that the former is a necessary condition for the latter must be defended more systematically. However, a full discussion of the issues involved will take us too far afield. In this paper, I am mainly interested in the nature of intransitive self-consciousness itself. I invite those readers who still feel that intransitive self-consciousness is not comfortably treated as a phenomenon of consciousness to take the argument of this paper as simply targeting the phenomenon of intransitive self-consciousness, whether or not it is related to consciousness. For a more extensive discussion of the role of self-awareness in consciousness, see U. Kriegel, 'Consciousness as Sensory Quality and as Implicit Self-Awareness,' *Phenomenology and the Cognitive Sciences* 2 (2003) 1-26.

6 In a philosophical world in which eliminative stances with regard to qualia, content, or propositional attitudes normally remain outside the mainstream, eliminativism about primordial forms of self-awareness has curiously been the norm. This is bemoaned extensively by contemporary German philosophers who follow the analytic philosophy of mind, especially members of the so-called 'Heidelberg School,' e.g., D. Henrich, 'Fichte's Original Insight,' D.R. Lachterman, *Contemporary German Philosophy* 1 (1982) 15-53; M. Frank, 'Mental Familiarity and Epistemic Self-Ascription,' *Common Knowledge* 4 (1995) 30-50; D. Sturma, 'Self-Consciousness and the Philosophy of Mind: A Kantian Reconsideration,' *Proceedings of the Eighth International Kant Congress* vol. 1 (Milwaukee, WI: Marquette University Press 1995). See also T. Kapitan, 'The Ubiquity of Self-Awareness,' *Grazer Philosophische Studien* 57 (1999) 17-44.

for any one to perceive, without perceiving, that he does perceive.⁷ It is a legitimate, and, to my mind, not altogether absurd, claim that these philosophers are all under an illusion. But in this paper I will assume that they are not.

What I want to discuss in this paper is the structure of consciousness in the sense of intransitive self-consciousness. Those who are uncertain of the existence of intransitive self-consciousness should therefore construe my thesis as a conditional: *if* there is such a thing as intransitive self-consciousness, *then* its structure is such-and-such.

II Two Accounts of Intransitive Self-Consciousness

Consciousness, in the sense of intransitive self-consciousness, is definitionally linked to self-awareness, because having a mental state self-consciously implies being aware of it. The link can be formulated as follows:

(SA) A mental state *M* of a subject *x* at a time *t* is conscious (i.e., intransitively self-conscious) only if *x* is aware of *M* at *t*.

Now, awareness of an object is ordinarily thought to employ some sort of mental representation of that object: if I am aware of a tree, I must be harboring a mental representation of the tree. Therefore, (SA) entails the following:

(SA₁) A mental state *M* of a subject *x* at a time *t* is conscious (i.e., intransitively self-conscious) only if *x* has at *t* a mental state *M**, such that *M** represents the occurrence of *M*.

That is, in order for *M* to be conscious, in the sense of being intransitively self-conscious, *x* must have at the same time a mental state which

7 Similar views are propounded by Descartes (see G. Rodis-Lewis, *Le probleme de l'inconscient et le cartesianisme* [Paris: Presses Universitaires de France 1950]; K. Wider, *The Bodily Nature of Consciousness: Sartre and Contemporary Philosophy of Mind* [Ithaca, NY: Cornell University Press 1997]), Cudworth (see U. Thiel, 'Cudworth and Seventeenth-Century Theories of Consciousness,' in S. Gaukroger, ed., *The Uses of Antiquity: The Scientific Revolution and the Classical Tradition* [Dordrecht: Kluwer 1991]), Leibniz (see R.J. Gennaro, 'Leibniz on Consciousness and Self-Consciousness,' in R.J. Gennaro and C. Huenemann, eds., *New Essays on the Rationalists* [New York: Oxford University Press 1999]), and Kant (see R.J. Gennaro, *Consciousness and Self-Consciousness* [Philadelphia: John Benjamin 1996]; D. Sturma; K. Wider), but the view goes all the way back to Aristotle (see V. Caston, 'Aristotle and Consciousness,' *Mind* 111 [2002] 751-815).

represents the occurrence of M .⁸ (From now on, I will not remind the reader that I am using ‘consciousness’ and ‘conscious’ specifically in the sense of intransitive self-consciousness.)

One thing SA_1 , as stated above, leaves open is whether M and M^* are to be construed as two numerically distinct mental states or as two aspects of one and the same state. When I have an auditory experience of a distant bagpipe, SA_1 says that the consciousness of the experience is a structure involving an awareness of the bagpipe sound and an awareness of this awareness of the bagpipe sound. But according to one interpretation, these are two distinct mental states, each with its own representational content, whereas according to a second interpretation, there is only one mental state involved, which carries a twofold representational content.

In other words, SA_1 is consistent both with (i) $M \neq M^*$ and with (ii) $M = M^*$ (if, that is, we restrict the labels ‘ M ’ and ‘ M^* ’ to the states themselves, in abstraction from their contents). Conjoined with (i), (SA_1) entails:

(SA_2) A mental state M of a subject x at a time t is conscious only if x has a mental state M^* , such that $M^* \neq M$, and M^* represents the occurrence of M .

But conjoined with (ii), (SA_1) entails:

(SA_3) A mental state M of a subject x at a time t is conscious only if M represents its own occurrence.

The difference between (SA_2) and (SA_3) is important, because according to (SA_3) consciousness is an intrinsic property of M , whereas according to (SA_2) it is a relational, extrinsic property, conferred on it from without. Moreover, (SA_3) introduces a special sort of intentional structure, whereby a mental state represents its very own occurrence — to which (SA_2) is not committed.⁹

8 It is not required by (SA_1) that M be the *only* thing that M^* represents. It may be one among several elements figuring in M^* 's content.

9 Some may take the question whether there are two distinct states involved or only one state to be a wholly arbitrary matter. On this view, what counts as one state or two states is up to us. And on this view, the difference between (SA_2) and (SA_3) is insignificant, if real at all. It seems to me, however, that what counts as one state or two states is not generally up to us. Thus, there is a psychologically real difference between believing that p and q and believing that p and that q . The former involves holding a single conjunctive belief while the latter involves holding a conjunction

Both these views have a venerable tradition behind them. (SA₂) is the view of current-day higher-order monitoring theorists of consciousness, such as Armstrong, Carruthers, Dennett, Lycan, Rosenthal, and Van Gulick.¹⁰ It may have also been the view of Locke, Kant, and other early Inner Sense theorists.¹¹ (SA₃) is more popular in the phenomenological tradition, where it was defended originally by Brentano,¹² and later by Brough, Frank, Gurwitsch, Henrich, Husserl, Natsoulas, Sartre, Smith, Sokolowski, Wider, Zahavi, and others.¹³ Recently, (SA₃) has enjoyed

of beliefs. That this is a real difference is manifest in the fact that a person may believe that *p* and that *q* without believing that *p* and *q*. (For a demonstration of this, see J.N. Williams, 'Inconsistency and Contradiction,' *Mind* 90 [1981] 600-2.) Moreover, as we will see below, (SA₂) leads to certain problems avoided by (SA₃). If so, the difference between them must be very real indeed.

- 10 D.M. Armstrong, *A Materialist Theory of the Mind* (New York: Humanities Press 1968), and 'What is Consciousness?' in N.J. Block, O. Flanagan, and G. Guzeldere, eds., *The Nature of Consciousness: Philosophical Debates* (Cambridge, MA: The MIT Press 1997); P. Carruthers, *Language, Thought, and Consciousness* (Cambridge: Cambridge University Press 1996); D. Dennett, *Content and Consciousness* (London: Routledge and Kegan Paul 1969); W.G. Lycan, 'Consciousness as Internal Monitoring,' *Philosophical Perspectives* 9 (1990) 1-14, and *Consciousness and Experience* (Cambridge, MA: The MIT Press 1996); D.M. Rosenthal, 'Two Concepts of Consciousness,' *Philosophical Studies* 94 (1986) 329-59, 'A Theory of Consciousness,' in Block et al., eds., *The Nature of Consciousness*, and *Consciousness and Mind* (Oxford: Oxford University Press forthcoming); R. Van Gulick, 'Inward and Upward — Reflection, Introspection, and Self-Awareness,' *Philosophical Topics* 28 (2001) 275-305.
- 11 Although this is debatable. Locke is commonly taken by current-day proponents of (SA₂) to have anticipated their view, but there is nothing in his writings to rule out (SA₃) and quite a bit to suggest it (see footnote 19 below). As for Kant, D.M. Rosenthal, 'Consciousness and the Mind,' *Iyyun* 51 (2002) 227-51, interprets him as a proponent of (SA₂), but Gennaro, *Consciousness and Self-Consciousness* interprets him as a proponent of (SA₃), or something very close to it. (According to Gennaro, Kant holds that M* is a non-conscious part of the same mental state M is a conscious part of.) Descartes, by contrast, was clearly more in line with the (SA₃) view (see Rodis-Lewis).
- 12 F. Brentano, *Psychology from Empirical Standpoint*, A.C. Rancurello, D.B. Terrell, and L.L. McAlister, trans., L.L. McAlister, ed. (London: Routledge and Kegan Paul 1973 [1874])
- 13 J.B. Brough, 'The Emergence of an Absolute Consciousness in Husserl's Early Writings on Time-Consciousness,' *Mind and World* 5 (1972) 298-326; M. Frank, 'Mental Familiarity'; A. Gurwitsch, *Marginal Consciousness* (Athens, OH: Ohio University Press 1985); D. Henrich, 'Fichte's Original Insight'; E. Husserl, *Phenomenology of Internal Time-Consciousness*, J.S. Churchill, trans., M. Heidegger, ed. (Bloomington, IN: Indiana University Press 1964); T. Natsoulas, 'The Case for Intrinsic

something of a revival in analytic circles as well, in the work of such writers as Carruthers, Caston, Gennaro, Levine, Thomasson, Van Gulick, and me.¹⁴

I want to defend the view of consciousness (in the sense of intransitive self-consciousness) captured in SA₃. To that end, I will explore an ancient line of argument in its favor. The argument itself suffers from a rather straightforward weakness, but it can be modified to carry more persuasion.

III The Aristotle-Brentano Line of Argument

Perhaps the best developed account of consciousness along the lines of (SA₃) is Brentano's. In his *Psychology from an Empirical Standpoint*, he states:¹⁵

Theory: I. An Introduction,' *Journal of Mind and Behavior* 17 (1996) 267-86; J.-P. Sartre, *La Transcendence de l'ego* (Paris: Vrin 1937) and *L'Être et le Néant* (Paris: Gallimard 1943); D.W. Smith, 'The Structure of (Self-)Consciousness,' *Topoi* 5 (1986) 149-56 and *The Circle of Acceptance* (Dordrecht: Kluwer 1989); R. Sokolowski, *Husserlian Meditations* (Evanston, IL: Northwestern University Press 1974); K. Wider, *The Bodily Nature of Consciousness*; D. Zahavi, 'Brentano and Husserl on Self-Awareness,' *Etudes Phenomenologiques* 27-8 (1998) 127-69 and *Self-Awareness and Alterity* (Evanston, IL: Northwestern University Press 1999)

14 P. Carruthers, *Phenomenal Consciousness* (Cambridge: Cambridge University Press 2000); V. Caston, 'Aristotle on Consciousness'; R. Gennaro, *Consciousness and Self-Consciousness*; J. Levine, *Purple Haze: The Puzzle of Consciousness* (New York: Oxford University Press 2001); A.L. Thomasson, 'After Brentano: A One-Level Theory of Consciousness,' *European Journal of Philosophy* 8 (2000) 190-209; R. Van Gulick, 'Inward and Upward'; U. Kriegel, 'Consciousness, Permanent Self-Awareness, and Higher-Order Monitoring,' *Dialogue* 41 (2002) 517-40 and 'Consciousness, Higher-Order Content, and the Individuation of Vehicles,' *Synthese* 134 (2003) 477-504. As you may notice, some philosophers who have defended (SA₂) earlier on have recently turned to (SA₃). Usually, they still call their views Higher-Order Monitoring — in particular, Carruthers and Van Gulick do — but the way I will use the terms here, Higher-Order Monitoring refers only to (SA₂)-style theories.

15 Brentano, 153-4. The view is first introduced in Section 7 of chapter II ('Inner Consciousness') in Book 2, which is entitled 'A Presentation and the Presentation of that Presentation are Given in One and the Same Act.' In this section, Brentano canvasses his conception of conscious experiences as self-representational.

[Every conscious act] includes within it a consciousness of itself. Therefore, every [conscious] act, no matter how simple, has a double object, a primary and a secondary object. The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard.

Brentano takes this view to have been Aristotle's. In *Metaphysics* 12.9, Aristotle writes that conscious 'knowing, perceiving, believing, and thinking are always of something else, but of themselves on the side' (1074b35-6).¹⁶

Brentano's 'master argument' against (SA₂) and in favor of (SA₃) is also borrowed from Aristotle, who writes this in the opening paragraph of *De Anima* III.2 (425b11-7):¹⁷

Since we perceive that we see and hear, it is necessarily either by means of seeing that one perceives that one sees or by another perception. But the same perception will be both of the seeing and of the color that underlies it, with the result that either two perceptions will be of the same thing, or it [sc. the perception] will be of itself. Further, if the perception of vision is a different perception, either this will proceed to infinity or some perception will be of itself; so that we ought to posit this in the first place.

This argument is developed and defended by Brentano throughout the second book of the *Psychology*. Its structure is quite straightforward. A conscious perception is itself perceived; so it is perceived either by itself, or by another perception; but it cannot be perceived by another perception, since this would lead to infinite regress of perceptions; therefore it must be perceived by itself.

Aristotle's formulation in terms of perception is limiting, because (i) the argument should extend to conscious thoughts (thoughts we have self-consciously), and (ii) there is no reason to premise the argument on the notion that our awareness of conscious states is necessarily perceptual (that is, the argument should be able to go through even if the

16 For more on the relationship between Aristotle and Brentano, see Caston. Caston notes that an interpretation of Aristotle along these lines is also to be found in a dissertation on the unity of mental life in Aristotelian philosophy, written under Brentano's supervision by one J. Herman Schell.

17 There is a long tradition of reading this passage in terms of perceptual *modalities*, or *senses*, rather than in terms of perceptual *states*, or *activities*. But that is not how Brentano reads it, and Caston argues in favor of Brentano's reading. My understanding is that the Greek word used by Aristotle, αἰσθησις, is ambiguous between sense and state. I am using here Caston's translation of the passage.

awareness is intellectual rather than perceptual). Accordingly, we can reformulate the argument as follows:

- (1) A mental state M of a subject x at a time t is conscious only if x is aware of M at t . (By definition of 'consciousness' in the relevant sense.)
- (2) If x is aware of M at t , then x 's awareness of M is either (i) part of M itself, or (ii) part of a distinct mental state M^* , where $M^* \neq M$. (Excluded middle.)
- (3) (ii) leads to infinite regress. Therefore,
- (4) (i) is the case; that is, x 's awareness of M is part of M itself. (2, 3.)
Therefore,
- (5) A mental state M of a subject x at a time t is conscious only if M constitutes (partly) awareness of itself. (1, 4.)

So construed, The Aristotle-Brentano argument proceeds by *reductio ad absurdum* of (SA₂). Evidently, the soundness of the argument depends on the truth of proposition (3). Let us look more closely, then, at the sub-argument for (3).

The gist of the sub-argument is this. To account for the fact that M is conscious, we posit M^* as a representation of M 's occurrence. If M^* was indeed a numerically distinct state, however, we would need to posit a *third* mental state, M^{**} , to account for the fact that M^* is conscious. And then we would need to posit a fourth mental state to account for the consciousness of M^{**} , and so on *ad infinitum*. The resulting account of consciousness is both absurd and empty: absurd, in that it entails that the occurrence of a single conscious state implies the occurrence of an infinity of them; and empty, in that the proposed explanation of consciousness appeals to the very same phenomenon it purports to explain (i.e., the explanandum shows up in the explanans).

The sub-argument, as it stands, is glaringly unsound. It presupposes that M^* , the awareness of M posited to explain M 's consciousness, must itself be a conscious state. The assumption derives from Brentano's Cartesian view that there are no unconscious mental states. To see its crucial role, let us reconstruct the argument a little more formally:

Let M_1 be a conscious state of a subject x at a time t .

- (1) For any M_i , if M_i is conscious, then x has at the same time a M_{i+1} , such that M_{i+1} represents M_i . (By definition of the relevant sense of consciousness.)

- (2) If M_1 is conscious, then x has at t an M_2 , such that M_2 represents M_1 . (1.)
- (3) M_1 is conscious. (Ex hypothesi.) Therefore,
- (4) x has at t an M_2 , such that M_2 represents M_1 . (2, 3.)
- (5) All mental states are conscious. Therefore,
- (6) M_2 is conscious. (5.)
- (7) If M_2 is conscious, then x has at t an M_3 , such that M_3 represents M_2 . (1.) Therefore,
- (8) x has at t an M_3 , such that M_3 represents M_2 . (6, 7.)
- (9) M_3 is conscious. (5.)
- ...

This procedure can be reiterated indefinitely, producing mental states M_2, M_3, M_4 , and so forth *ad infinitum*. The way the argument is driven home can be captured as follows:

- (10) x has at t mental states M_2, M_3, M_4, \dots such that each M_i is represented by M_{i+1} . (By infinite repetition of 1-9.)
- (11) For any M_i and M_{i+1} , either $M_i = M_{i+1}$ or $M \neq M_{i+1}$. (Excluded middle.)
- (12) If for every M_i and M_{i+1} , $M \neq M_{i+1}$, then x has at t infinitely many mental states. (10, 11.)
- (13) x does not have at t infinitely many mental states. (Empirical.) Therefore,
- (14) For some M_i and M_{i+1} , $M_i = M_{i+1}$. (12, 13.) But,
- (15) For any M_i, M_{i+1} , and M_{i+2} , there is no more reason to suppose that $M_{i+1} = M_{i+2}$ than there is to suppose that $M_i = M_{i+1}$. Therefore, we may as well suppose that,
- (16) $M_1 = M_2$.

The crucial premise in this reconstructed argument is (5). The other premises are (1), (13), and (15), which are all hard to deny.¹⁸ So the argument falters mainly on (5). The Cartesian notion that all mental states are conscious is plainly false, as one of Brentano's students — a certain Sigmund Freud — has compellingly argued.¹⁹ If so, the proponent of (SA₂) does not have to admit that M₂ is also conscious, and the infinite regress is avoided. Without the infinite regress, proposition (3) in the overall argument remains unjustified.

This is precisely the line taken by present-day Higher-Order Monitoring theorists. In their view, M₁ is conscious in virtue of being represented by M₂, which is numerically different from M₁, but M₂ is a *non-conscious* representation of M₁. Since M₂ is non-conscious, it does not call for positing an M₃ which would represent it.²⁰

IV Another Version

In discussing Brentano's reductio, we claimed that what he failed to take into account was the existence of non-conscious mental states. But perhaps this is because he worked with the assumption that only *occurrent* mental states have psychological reality.

It is in fact quite likely that he worked with such an assumption. Presented with the notion of Freudian or other unconscious states, he

18 (1) is definitional; (13) is empirically true; and (15) is a methodological point I see no justification for rejecting.

19 As a historical side note, it is interesting to note that Freud appears to have held Brentano's view about conscious states — that they represent their own occurrence within the subject — but did not couple it with the Cartesian notion that all mental states are conscious. His position was that not all mental states were conscious, but those which were represented themselves (see T. Natsoulas, 'Freud and Consciousness: I. Intrinsic Consciousness,' *Psychoanalysis and Contemporary Thought* 7 [1984] 195-232). This position is of course coherent, and in my opinion is in fact true, but it undermines Brentano's specific argument for his conception of conscious states as self-representational.

20 Leibniz may have anticipated this line of reasoning. Leibniz is perhaps the first philosopher to have recognized the existence of non-conscious states ('small thoughts'), and he offers the following way out of the infinite regress: 'It is impossible that we should always reflect explicitly on all our thoughts; [otherwise] the mind would reflect on each reflection *ad infinitum*.... It must be that ... eventually some thought is allowed to occur without being thought about; otherwise I would dwell forever on the same thing' (quoted in Gennaro, 'Leibniz on Consciousness,' 355-6).

would probably deny that they are *psychologically real* in the way occurrent conscious states are. We can call such psychological factors ‘latent mental states,’ or ‘dispositional mental states,’ if we want to, but this is not to be understood as implying that these are real, concrete items in one’s mind.

Let us set aside, for now, the question whether such an assumption would be warranted. The assumption has implications for how we construe Brentano’s argument. As reconstructed above, the argument depends on the implausible principle that all mental states are conscious. But it would be more charitable to read Brentano as employing the more plausible principle that all *occurrent* mental states are conscious.

If this is indeed the principle Brentano had in mind, the *reductio* should be reconstructed rather as follows:

Let M_1 be a conscious state of a subject x at a time t .

- (1) For any M_i , if M_i is conscious, then x has at the same time an M_{i+1} , such that M_{i+1} is an occurrent representation of M_i .
- (2) If M_1 is conscious, then x has at t an M_2 , such that M_2 is an occurrent representation of M_1 . (1.)
- (3) M_1 is conscious. (*Ex hypothesi.*) Therefore,
- (4) x has at t an M_2 , such that M_2 is an occurrent representation of M_1 . (2, 3.)
- (5) All occurrent mental states are conscious. Therefore,
- (6) M_2 is conscious. (5.)
- (7) If M_2 is conscious, then x has at t an M_3 , such that M_3 is an occurrent representation of M_2 . (1.)

...

And so on and so forth. In this version, Brentano’s *reductio* looks much more plausible than in the previous one.²¹

The *reductio* is still problematic, though. While the principle that all occurrent mental states are conscious (i.e., premise (5) in the present version of the *reductio*) is more palatable than the principle that all mental states are conscious (i.e., premise (5) in the previous version),

21 Caston argues that this is precisely how Aristotle meant his argument. The word Aristotle uses is not ‘states,’ but ‘activities,’ which suggests an occurrent reading.

there is still good reason to think it false. For ‘conscious’ and ‘occurrent’ are not coextensive: there are such things as non-conscious occurrent mental states.

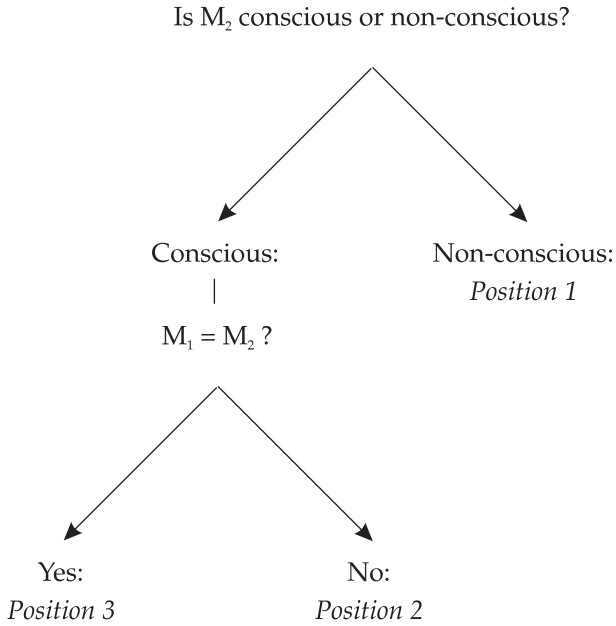
Consider the following case.²² You and your friend are stepping into the car while engaged in a heated conversation. As your friend unfolds his take on last year’s disputed elections, he reaches absent-mindedly into his jacket pocket, in search of the car keys. Your friend is too steeped in his passionate sermon to take any notice of what he is doing, but surely he digs in his pocket for a reason. He digs in his pocket because he thinks the keys are there. But the thought that the keys are in his pocket is not merely dispositional; it is occurrent. That is, he has an occurrent thought that the keys are in his pocket. Yet the thought is not conscious in the relevant sense, since he does not have it self-consciously. Your friend’s thought is a non-conscious occurrent mental state. Thus there are non-conscious occurrent mental states — contrary to (5) above.²³

V A Twist

Brentano’s argument is often dismissed as worthless, but as we just saw, with a little charity it becomes at least somewhat plausible. In fact, the argument will not appear so utterly insignificant if it is reconstructed as an *argument by elimination*. So construed, the argument could run as follows. (SA₁) states that if M₁ is conscious, then x must have an M₂, such that M₂ represents the occurrence of M₁. But (SA₁) is silent on whether M₂ is itself conscious or non-conscious. If M₂ is non-conscious, then necessarily M₁ ≠ M₂, since M₁ is conscious. If M₂ is conscious, however, then both M₁ = M₂ and M₁ ≠ M₂ are open possibilities. So we have three possible positions consistent with SA₁:

22 Adapted from N. Malcolm, ‘Thoughtless Brutes,’ in D.M. Rosenthal, ed., *The Nature of Mind* (New York: Oxford University Press 1991).

23 Another, less disturbing problem is that while premise (1) in the previous version of the argument was more or less definitionally true, premise (1) in the present version is not. Thus, according to proponents of the Dispositional Higher-Order Monitoring theory (e.g., Carruthers, *Language, Thought, and Consciousness*), what makes a mental state conscious is not that it is represented by an occurrent mental state, but that it is represented by a dispositional mental state. That is, they reject premise (1) of the current version of the argument. Their view may not be very plausible, but it is nowise ruled out by the present version of Brentano’s argument.



The three positions can be stated as follows:

Position 1: M_1 is conscious only if x has an M_2 , such that $M_2 \neq M_1$, and M_2 is a *non-conscious* representation of M_1 .

Position 2: M_1 is conscious only if x has an M_2 , such that $M_2 \neq M_1$, and M_2 is a *conscious* representation of M_1 .

Position 3: M_1 is conscious only if M_1 is a conscious representation of *itself* (perhaps among other things).

Positions 1 and 2 are two versions of (SA_2) , while Position 3 is just (SA_3) . Position 1 is the position of current-day Higher-Order Monitoring theorists. Position 2 is perhaps what Locke had in mind.²⁴ And Position 3 is

²⁴ This must be the case if Locke is to be interpreted as a proponent of (SA_2) at all. I am not sure that he was, but in any event it is clear that he did not countenance non-conscious states: 'Whilst [the soul] thinks and perceives ... it must necessarily

the position for which Brentano argues. His argument can be construed as proceeding by elimination, first of Position 1 and then of Position 2. Position 1 is eliminated on the grounds that there are no non-conscious (occurrent) states, so M_2 cannot be non-conscious, while position 2 is eliminated on the grounds that it leads to infinite regress. Now, while the elimination of Position 1 is premised on the indefensible notion that all (occurrent) mental states are conscious, and is thus unsound, the elimination of Position 2 is sound: M 's being conscious cannot be explained in terms of its being represented by a distinct *conscious* state, because such an explanation would lead to infinite regress.

This is, perhaps, why nobody today holds anything like Position 2. Another reason, though, may be that Position 2 appears inconsistent with the unity of consciousness. In one sense of the phrase, the unity of consciousness refers to the fact that the contents of our states of consciousness are cohesive and unified.²⁵ However, if our consciousness could be sometimes divided into two distinct states (such that the subject would be in two separate conscious states in parallel), that would introduce a *disunity* into the content of our state of consciousness. Thus the unity of consciousness entails that normally we have no more than one conscious state at a time.

In any event, the fact that Brentano's argument is weak only in its elimination of Position 1 suggests that an alternative elimination of it may pave the way to a viable argument for (SA_3). That is, if M_2 's being conscious can be defended on new grounds, other than the principle that all mental states are conscious, or that all occurrent mental states are conscious, the emerging argument may be sound.

This argument by elimination can be completed, then, by plugging into it any of the accepted arguments against the position of current-day Higher-Order Monitoring theory. If the position of the Higher-Order Monitoring (henceforth, HOM) theorist can be rejected on plausible grounds, the overall argument for (SA_3) will be correspondingly plausible.

Arguments against the HOM theory abound in the literature. Aquila, Byrne, Carruthers, Dretske, Goldman, Guzeldere, Levine, Moran, Nat-

be conscious of its own perceptions' (*Essay* II, i, 12). This means that Locke was either oblivious to the infinite regress consequent upon holding Position 2, or held rather something like Position 3 and is wrongly appropriated by current-day Higher-Order Monitoring theorists.

25 In another sense, it refers to the fact that all or many of the subject's conscious states are united in a single personal consciousness. But that is not the sense I am interested in here.

soulas, Neander, Rey, and Shoemaker have all argued against one or another version of the idea.²⁶ Perhaps the most troublesome difficulty emerging from this literature is the problem of targetless higher-order states (developed by Byrne and Neander).²⁷ Let me gloss over it quickly; I will then present what I take to be the deep problem with, and the source of the intuitive implausibility of, the HOM theory.

According to the HOM theory, x 's state M_1 is conscious only if x is also in state M_2 , which represents the occurrence of M_1 . Suppose that on some occasion x is in M_2 , which represents the occurrence of M_1 , without actually being in M_1 . That is, suppose that M_2 misrepresents the occurrence of M_1 . There are, in fact, two ways for M_2 to be misrepresentational. One is to represent M_1 to be so-and-so when in reality M_1 is not so-and-so. The other is to represent M_1 to be so-and-so when in reality M_1 does not even exist. My interest here is in the second form of higher-order misrepresentation. In such circumstances, is x in a conscious state or not? The HOM theorist must say, on the one hand, that x is *not* in a conscious state, but on the other hand, that it *seems* to x as though she is in a conscious state. She must say that x is not in a conscious state, because neither M_2 nor M_1 can be a conscious state x is in. M_2 is, according to HOM theory, a non-conscious state, and M_1 is, *ex hypothesi*, non-existent,

26 R. Aquila, 'Consciousness and Higher-Order Thoughts: Two Objections,' *American Philosophical Quarterly* 27 (1990) 81-7; A. Byrne, 'Some Like it HOT: Consciousness and Higher-Order Thoughts,' *Philosophical Studies* 86 (1997) 103-29; P. Carruthers, *Phenomenal Consciousness*; F.I. Dretske, 'Conscious Experience,' *Mind* 102 (1993) 263-83; A. Goldman, 'Consciousness, Folk Psychology, and Cognitive Science,' *Consciousness and Cognition* 2 (1993) 364-83; G. Guzeldere, 'Is Consciousness the Perception of What Passes in One's Own Mind?' in *Conscious Experience*, T. Metzinger, ed. (Padborn: Schoeningh-Verlag 1995); J. Levine, *Purple Haze*; R. Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press 2001); T. Natsoulas, 'What Is Wrong with Appendage Theory of Consciousness?' *Philosophical Psychology* 6 (1993) 137-54; K. Neander, 'The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness,' *Philosophical Perspectives* 12 (1998) 411-34; G. Rey, 'A Question about Consciousness,' in *Perspectives on Mind*, H. Otto and J. Tueidio, eds., (Norwell: Kluwer 1988); S. Shoemaker, 'Self-Knowledge and "Inner Sense." Lecture II: The Broad Perceptual Model,' *Philosophy and Phenomenological Research* 54 (1994) 271-90.

27 The Higher-Order Monitoring theory comes in two varieties, depending on how the second-order state is construed. If it is construed as a perception-like state, we have a Higher-Order Perception theory; if it is construed as an intellectual thought, we have a Higher-Order Thought theory. Byrne presents the problem of targetless higher-order states for (Rosenthal's) Higher-Order Thought theory, whereas Neander presents it for (Lycan's) Higher-Order Perception theory. Levine develops the argument in a generic way, against both varieties of Higher-Order Monitoring theory.

that is, a state x is not in. (If M_1 is not a state x is in, it can hardly be a *conscious* state x is in.) At the same time, the HOM theorist must say that it *seems* to x as though she is in a conscious state, because, through M_2 , x is aware of herself as being in a conscious state.

That it will indeed seem to x as though she is in a conscious state, even though she is not, is claimed explicitly by Rosenthal ('A Theory of Consciousness,' 744; italics mine):

Strictly speaking, having a HOT [higher-order thought] cannot of course result in a mental state's being conscious if that mental state does not even exist.... Still, a case in which one has a HOT along with a mental state it is about may be *subjectively indistinguishable* from a case in which the HOT occurs but not the mental state. If so, folk psychology would count both as cases of conscious states.²⁸

Folk Psychology would, indeed, but HOM theory would not — it could not. But this is not one of the marginal inaccuracies of folk psychology, that we would be better off without. It is quite preposterous to suppose that a person may be under the impression of being conscious when in fact she is unconscious. This is a major counter-intuitive consequence of HOM theory, even if its proponents consistently try to downplay it.

Observe that the same difficulty does *not* attend the view that M_1 is conscious in virtue of representing its own occurrence. In circumstances in which M_1 does not occur, M_1 cannot misrepresent that it does occur. If M_1 does not exist, then it does not represent anything. And if it does not represent anything, *a fortiori* it does not represent its own occurrence. Thus for M_1 to represent its own occurrence, indeed for M_1 to represent anything (its own occurrence included), M_1 must exist. It is therefore impossible for M_1 to misrepresent its own occurrence.²⁹

Beyond the various more or less technical difficulties facing the Higher-Order Monitoring theory, I think there is a quite straightforward, and principled, reason many have found it so unappealing; namely, that the awareness of our conscious states is something we *experience*. The

28 Quoting this very passage, Levine (109, n.24) comments: 'But doesn't this give the game away?' To all appearances, it does.

29 There is no Cartesian voodoo involved in this form of immunity to misrepresentation. It is simply an artifact of the fact that M_1 's occurrence is a condition of M_1 's representation of its occurrence. That is, the obtaining of the state of affairs M_1 purports to represent is a condition of M_1 's representation of this state of affairs, and this is why M_1 cannot misrepresent its own occurrence. So M_1 can fail to occur or it can represent its own occurrence while occurring, but it cannot represent its own occurrence while failing to occur. Recall, however, that M_1 can perfectly well misrepresent anything else it may represent (other than its own occurrence).

intuitive attraction of the notion that consciousness is a monitoring device derives from our first-person feeling that we are aware of our conscious states. But once the theory withdraws to the position that the awareness of our conscious states is a non-conscious representation of them, it becomes evident that the awareness posited by the theory is something different from the phenomenon that lent the theory its initial plausibility.

Consider, for instance, your conscious perception of the paper before your eyes. Since you have this perception self-consciously (i.e., you self-consciously perceive the paper), you have some dim, background awareness of it. But this awareness is something of which you have first-person experience. So it would be quite strange to consider it a non-conscious state, given that you experience it in real time.

This line of reasoning can be represented in the following straightforward argument against Position 1:

- (1) The occurrence of M_2 is experienced by x .
- (2) Experienced mental states are conscious. Therefore,
- (3) M_2 is conscious.

The argument's conclusion rules out Position 1, which says that M_2 is non-conscious. Once Position 1 is thus falsified, the argument by elimination is completed. For as we already saw, Position 2 is false too (it leading to infinite regress). Therefore, the state M_2 must be identical to the state M_1 : your awareness of the paper and your awareness of your awareness of the paper are two aspect of one self-identical state.

In the remainder of this section, I defend premises (1) and (2) in this argument. In the next section, I will consider objections to the overall argument by elimination.

Premise (1) states that the awareness of M_1 is something of which we have first-person experience. Now, recall that M_2 is the awareness of M_1 in virtue of which M_1 is intransitively self-conscious: if you self-consciously perceive the paper before you, you are minimally aware of perceiving the paper, and this awareness is M_2 . This awareness is not something we are inclined to posit on theoretical or explanatory grounds. If we are inclined to admit such awareness at all, it is on first-personal, experiential grounds. Conversely, those who insist that they do not find in their experience anything like an awareness of their conscious perceptions and thoughts probably deny the very existence of intransitive self-consciousness. What is difficult to imagine is someone who accepts the existence of intransitive self-consciousness, but insists she has no first-personal grounds for doing so — someone who thinks

that the reasons for positing an awareness of our conscious states is purely theoretical or third-personal.

This is what Higher-Order Monitoring theorists do. They would have us believe that the only reason to recognize the existence of an awareness of our intransitively self-conscious states is the theoretical or explanatory benefits consequent upon doing so. But what exactly is the explanatory role that M_2 — the awareness of M_1 — is called on to fulfil? What kinds of data are explained only by the positing of a non-conscious representation of M_1 ? No, the reason we recognize M_2 's existence is that we experience it. When I offered the phenomenological characterization of the non-reflective mode of self-awareness humming in the background of our mind, you knew exactly what I was talking about. We did not posit this self-awareness on *purely theoretical grounds*. Rather, we *pointed* to it as a familiar element in the human mental life, an element every conscious person is permanently experiencing. This self-awareness is not experienced at the focal center of our conscious awareness, to be sure — after all, it is humming, as I said, in the *background* of conscious awareness — but it is a constant element in the fringe of our experience.

The second premise of the sub-argument against Position 1 is that mental phenomena of which we have first-person experience are conscious. This appears to be something of a conceptual truth: conscious states are those states the subject *experiences*. The general principle operative in *this* elimination of Position 1, then, is not the principle that all mental states are conscious (which is clearly false), nor the principle that all *occurrent* mental states are conscious (which is also false), but the general principle that all *experienced* mental states are conscious (which is more or less conceptually true).

Another way to put this argument against Position 1 is as follows. Conscious states we are presently in are states we have first-person knowledge of being presently in. But if the HOM theory was correct, it would have to be third-person knowledge. For according to HOM theory, to say that a mental state M is conscious is to say that its subject harbors a non-conscious representation of M . So for the subject to know that she is in conscious state M is for her to know that she harbors a non-conscious representation of M , that is, to know that she harbors a certain non-conscious state (the representation of M). And knowledge that one harbors a non-conscious state is third-person knowledge, not first-person knowledge. Therefore, if HOM theory was correct, our knowledge of the conscious states we are presently in would be third-person knowledge. In this form, the argument would unfold thus:

- (1) x 's knowledge that she is in M_1 is first-person knowledge;

- (2) If M_1 was conscious only if x had an M_2 , where $M_2 \neq M_1$ and M_2 is a *non-conscious* representation of M_1 (that is, if Position 1 was true), then x 's knowledge that she is in M_1 would *not* be first-person knowledge; therefore,
- (3) It is not the case that M_1 is conscious only if x has an M_2 , such that $M_2 \neq M_1$ and M_2 is a *non-conscious* representation of M_1 (that is, Position 1 is false).

Thus your knowledge that you now have a conscious experience of this page is first-person knowledge: you know this in a very direct, non-inferential way. But if HOM theory was right, then to know that you now have a conscious experience of this page would be to know that you are in a certain non-conscious mental state, which knowledge you would infer from certain cues, or on the basis of certain evidence. This may or may not be an easy thing to do, but it is clearly not what you actually do when you acquire the knowledge that you now have a conscious experience of this page.³⁰

This argument against Position 1 is intended to capture the principled reason most of us remain skeptical about the Higher-Order Monitoring theory. But recall that any argument against this theory is welcome as part of the overall argument by elimination. The overall argument for Position 3 proceeds as follows:

- (1) A mental state M of a subject x at a time t is conscious (i.e., intransitively self-conscious) only if x is aware of M at t . (By definition of 'consciousness' in the relevant sense.)
- (2) If x is aware of M at t , then x 's awareness of M is either (i) part of M itself, or (ii) part of a distinct conscious state M^* , or (iii) part of a distinct non-conscious state M^* . (Exhaustive list.)
- (3) (ii) leads to infinite regress.

30 Goldman's ('Consciousness, Folk Psychology') argument against the HOM theory unfolds along these lines. It is not essential to this argument that it be coupled with an account of the difference between first-person knowledge and third-person knowledge. All is required is a commitment to the existence of such a distinction. Some philosophers may feel that the very distinction is an affront to a naturalist conception of knowledge. But this is clearly misguided: if there is a distinction between first-person and third-person knowledge, it must be possible to account for it in naturalist terms. For an attempt to explain the distinction in naturalist terms, see, e.g., F.I. Dretske, *Naturalizing the Mind* (Cambridge, MA: The MIT Press 1995), J. Fernandez, 'Privileged Access Naturalized,' *Philosophical Quarterly* 53 (2003) 352-72.

- (4) (iii) is incompatible with the fact that x 's awareness of M is experienced (and faces a number of other difficulties). Therefore,
- (5) (i) is the case; that is, x 's awareness of M is part of M itself. (2, 3, 4.) Therefore,
- (6) A mental state M of a subject x at a time t is conscious (i.e., intransitively self-conscious) only if M constitutes (partly) awareness of itself. (1, 5.)³¹

This argument is very similar to (and, after all, based on) the Aristotle-Brentano argument outlined in §II. But unlike that argument, it appears to be sound. If so, a mental state is intransitively self-conscious when, and only when, it partly constitutes an awareness of itself. Thus, my perception of the laptop before me is intransitively self-conscious (i.e., I self-consciously perceive the laptop) because the perception wraps up together, in a single mental act, an awareness of the laptop and an awareness of itself.

31 A fuller formulation of the argument would go as follows:

- (1) A mental state M of a subject x at a time t is conscious (i.e., intransitively self-conscious) only if x is aware of M at t . (SA.)
- (2) An awareness of a thing T requires a mental representation of T . (Assumption.)
- (3) A mental state M of a subject x at a time t is conscious (i.e., intransitively self-conscious) only if x has at t a mental state M^* , such that M^* represents the occurrence of M . (SA₁.)
- (4) Either (i) $M^* = M$ or (ii) $M^* \neq M$. (Excluded Middle.)
- (5) If (ii), then a mental state M of a subject x at a time t is conscious (i.e., intransitively self-conscious) only if either (ii.i) x has at t a mental state M^* , $M^* \neq M$, such that M^* is a *non-conscious* representation of the occurrence of M , or (ii.ii) x has at t a mental state M^* , $M^* \neq M$, such that M^* is a *conscious* representation of the occurrence of M . (Excluded middle.)
- (6) The occurrence of M^* is experienced.
- (7) Experienced mental states are conscious. Therefore,
- (8) M^* is a conscious state. (6, 7.) Therefore,
- (9) Not-(ii.i). (8.)
- (10) (ii.ii) leads to infinite regress. Therefore (by reductio),
- (11) Not-(ii.ii). (10.) Therefore,
- (12) A mental state M of a subject x at a time t is conscious (i.e., intransitively self-conscious) neither only if (ii.i) x has at t a mental state M^* , $M^* \neq M$, such that M^* is a *non-conscious* representation of the occurrence of M , nor only if (ii.ii) x has at t a mental state M^* , $M^* \neq M$, such that M^* is a *conscious* representation of the occurrence of M . (9, 11.) Therefore,
- (13) Not-(ii). (5, 12.) Therefore,
- (14) (i) is the case; that is, M_1 is conscious (i.e., intransitively self-conscious), only if M is a conscious representation of *itself*. (4, 13.)

I conclude that a mental state is intransitively self-conscious only if it represents its own occurrence. That is, self-representation is a necessary condition for intransitive self-consciousness. Recall, now, that in §I I suggested that intransitive self-consciousness is a necessary condition for phenomenal consciousness: a mental state is phenomenally conscious only if it is intransitively self-conscious. It follows from these two claims that a mental state is phenomenally conscious only if it represents its own occurrence, that is, that self-representation is a necessary condition for phenomenal consciousness.

VI Objections and Replies

I close with a discussion of seven different objections. The first two are objections against the *view* argued for, i.e., Position 3, or (SA₃). The five subsequent objections target the *argument* for the view.

First objection. Position 3 leads to an equally vicious regress as Position 2—not a regress of conscious states, but a regress of their representational contents. Even if M_2 is the same *state* as M_1 , it is nonetheless a different *content*. For instance, where M_1 is a conscious experience of the blue sky, the M_1 -content is the blue sky, but, according to Position 3, the M_2 -content is M_1 itself, the *representation* of the blue sky. So given that M_2 is conscious as well, there would have to be an M_3 whose content is M_2 , and so on and so forth. And even though M_1 , M_2 , M_3 , etc. are all one and the same conscious *state*, that single state will have to carry infinitely many contents. If so, Position 3 is susceptible to the same kind of *reductio* as Position 2.

This objection would be valid if what each M_{i+1} represented was the *content* of M_i . But according to Position 3, what each represents is the *state* M_i . So there is no infinite regress of contents, since the M_2 -content, M_3 -content, M_4 -content, etc. collapse into one: they all represent the ground-level state M_1 . Consider again the experience of the blue sky. The M_1 -content is the blue sky, the M_2 -content is state M_1 , the M_3 -content is state M_2 , the M_4 -content is state M_3 , etc. But since $M_1 = M_2 = M_3$, the M_2 -content is identical with the M_3 -content, which is identical with the M_4 -content, and so on.

Let me examine this objection a little closer, because it has been presented to me many times in connection with the Brentanian view of consciousness.³² To get a better grip on the objection, let us use the

32 It is also quite common in the phenomenological tradition to accuse Brentano of falling prey to the same infinite regress he used in order to argue for his view. This is one of the main tenets of the Heidelberg School (see M. Frank), and is often emphasized by Zahavi (see 'Brentano and Husserl').

notation $C(M_i)$ to represent the content of mental state M_i . According to Position 3, when I self-consciously think that the almond trees are blooming again, I have a thought M_1 with two contents: the primary content is the proposition \langle The almond trees are blooming again \rangle , whereas the secondary content is something like the proposition \langle I am herewith thinking that the almond trees are blooming again \rangle . In general,

M_i is a conscious thought that p only if

(a) $C_1(M_i) = \langle p \rangle$

(b) $C_2(M_i) \cong \langle$ I am herewith thinking that $p \rangle$

This seems to presuppose a general principle regarding the content of conscious states, a principle we may call the 'secondary content principle' and state as follows:

(SCP) For a conscious thought M_i , if $C_1(M_i) = \langle p \rangle$, then $C_2(M_i) \cong \langle$ I am herewith thinking that $p \rangle$.

The objection can be understood as follows, then. Since \langle I am herewith thinking that $p \rangle$ is a content of M_i , (SCP) will require that M_i have a further content, $C_3(M_i) = \langle$ I am herewith thinking that I am herewith thinking that $p \rangle$. And this will require a further content $C_4(M_i)$, and we are off with the regress again.

But the objection is fallacious. The content principle it would need to derive the existence of a third content $C_3(M_i)$ is this:

For a conscious thought M_i , if $C_i(M_i) = \langle p \rangle$, then $C_{i+1}(M_i) \cong \langle$ I am herewith thinking that $p \rangle$.

This principle, however, is different from (SCP). (SCP) provides for the introduction of a second-order content depending on the first-order content of the conscious state, but the principle employed by the objector provides for the introduction of a higher-order content given *any* lower-order content, not just the first-order content. This more general principle is something the Brentanian approach to consciousness, captured in Position 3, is not committed to and in fact rejects (as it must).

Second objection. The Brentanian approach of Position 3 makes use of the notion of a self-representing mental state. But this notion is quite mysterious. What is involved in a mental state representing its very own occurrence? In particular, to the extent that we hope to *naturalize* consciousness, we would want to know how a physical system such as the brain can enter self-representing states and how self-representation can be realized in neural substrate.

My reply is twofold. First, this is indeed a challenge for the proponent of Position 3, a challenge that will have to be reckoned with eventually.

Some work has already started on trying to understand how self-representational states can be realized in a purely physical system. Thus, Carruthers (*Phenomenal Consciousness*), Van Gulick ('Inward and Upward'), and I ('Consciousness, Permanent Self-Awareness,' 'Consciousness, Higher-Order Content') have recently offered naturalistic models of consciousness along the lines of Position 3.³³

But beyond that, it should come as no surprise that an account of consciousness includes a *prima facie* mysterious element. For consciousness, while perhaps not *ultimately* mysterious, is surely *prima facie* mysterious, and its *prima facie* mysteriousness *should* be reflected in a *prima facie* mysterious element in the account of it. Seen in this light, Position 3 is interpreted as a suggestion about what it is that makes consciousness *prima facie* mysterious. The suggestion is that it is the self-representational character of conscious states that is the source of the *prima facie* mystery. What a demystification of consciousness would have to involve, according to this suggestion, is the demystification of self-representation. Far from being a weakness of Position 3, its mysterious allure is actually an advantage. Normally, an account of consciousness that does not provide for any sense of *prima facie* mystery is greeted with suspicion.³⁴ This is not the kind of problem Position 3 faces.

Third objection. My usage of the phrase 'to experience a mental state,' vital though it is for the sub-argument against Position 1, is unclear and perhaps even muddled. What does it mean to say that a mental state is *experienced*? Ordinarily, when we say that a thing T is experienced by x,

33 Furthermore, some proponents of (SA₂) accept the existence of self-representing states, but just do not think that *all* conscious states are such (e.g., D.M. Rosenthal, 'Higher-Order Thoughts and the Appendage Theory of Consciousness,' *Philosophical Psychology* 6 [1993] 155-66). This amounts, however, to an admission that there is nothing inherently mysterious about mental states that carry self-representational content.

34 D.J. Chalmers, 'Facing Up to the Problem of Consciousness,' *Journal of Consciousness Studies* 2 (1995) 200-19, gives a succinct expression of this suspicion: 'It is common to see a paper on consciousness begin with an invocation of the mystery of consciousness, noting the strange intangibility and ineffability of subjectivity, and worrying that so far we have no theory of the phenomenon. Here the topic is clearly the hard problem — the problem of experience. In the second half of the paper, the tone becomes more optimistic, and the author's own theory of consciousness is outlined. Upon examination, this theory turns out to be a theory of the more straightforward phenomena — of reportability, of introspective access, or whatever. At the close, the author declares that consciousness has turned out to be tractable after all, but the reader is left feeling like a victim of a bait-and-switch. The hard problem remains untouched' (211).

we mean that T is *the object* (or content) of an experiential state of *x*'s, not that T is itself the experiential state. If so, the fact that the awareness of M_1 is experienced means that it is the object of an experience, not that it is itself an experiential state.

To clarify the matter, let us bring up a distinction between two uses of such verbs as 'to experience' and 'to kick'.³⁵ Consider first verbs like 'to smile' and 'to dance.' There is a perfectly grammatical sense in which we can say that a person smiled a smile (e.g., a wide smile) or danced a dance (e.g., a tango). More than that, one cannot smile anything *but* a smile nor dance anything *but* a dance. Experiencing and kicking are a little different: it is *also* grammatical to say that one experiences something other than the experience, or kicks something other than a kick. Thus, one can experience a tree or kick a stray hound. The thing to notice, though, is that 'experiencing' and 'kicking' still preserve the sort of usage we found in 'smiling' and 'dancing.' One can be said to experience an experience (e.g., a headache) or to kick a kick (e.g., a scissors) in the same sense one can be said to smile a smile or dance a dance.

With this distinction at our disposal, we can get clearer on the sense in which the awareness of M_1 can be said to be experienced. This is the usage we noted first, the usage 'experiencing' shares with 'smiling' and 'dancing.' In this usage, to say that the awareness of M_1 is experienced is precisely to say that it is an experiential state, just as to say that a certain tango is danced is to say that it is a dance. This is the usage of 'experiencing' that is relevant to the claim that the awareness of M_1 is experienced by us.

Fourth Objection. The amended argument fallaciously infers how things are from how they *seem* to be, or how we *experience* them to be — a mode of argument all too popular in the phenomenological tradition.

This objection misfires. It is important not to confuse the above argument against Position 1 with a different argument, which one often encounters in the phenomenological literature. That other argument claims that our awareness of our conscious states *seems* to be intrinsic to these states, and therefore it likely *is* intrinsic to them (see Smith, *The Circle of Acquaintance* and T. Natsoulas, 'The Case for Intrinsic Theory: IV,' *Journal of Mind and Behavior* 20 (1999) 257-76; and Rosenthal, 'Two Concepts' for a criticism). The argument offered above, by contrast, is not an inference from how M_2 seems to how M_2 likely is. Rather, it is an inference from the fact that M_2 seems *at all* to the conclusion that M_2 is

35 The distinction is noted by E. Sosa, 'Experience and Intentionality,' *Philosophical Topics* 14 (1986) 67-83, at 73.

conscious.³⁶ Whatever the merits of the phenomenological argument for Position 3, the present objection is an *ignoratio elenchi*.

Fifth Objection. Now that Position 1 is soundly eliminated, the elimination of Position 2 has become unsound. For the infinite regress from Brentano's original argument cannot be reproduced in the amended argument: while we may grant it is empirically true that M_2 is experienced, it is not true that M_3 is experienced, let alone that M_4 is experienced. Thus Position 2 can be played out as follows: the consciousness of M_1 is to be explained by positing an experienced, hence conscious, mental state M_2 , which represents the occurrence of M_1 , and an inexperienced, hence non-conscious, mental state M_3 , which represents the occurrence of M_2 . The emerging position is this:

Position 2': M_1 is conscious only if x has an M_2 and an M_3 , such that

- (a) $M_1 \neq M_2$, $M_1 \neq M_3$, and $M_2 \neq M_3$;
- (b) M_2 is a conscious representation of M_1 ; and
- (c) M_3 is a non-conscious representation of M_2 .

Thus this version of Position 2 avoids the infinite regress.

Position 2' does not avoid the infinite regress, however. According to Position 2', M_1 is conscious in virtue of x having two mental states, one a conscious representation of M_1 and one a non-conscious representation of that conscious representation. Now, call the first of these two states M^* . The fact that M^* is conscious entails, by this very analysis, that x has a *further* pair of mental states, one a conscious representation of M^* and one a non-conscious representation of that conscious representation of M^* . Now, call the first of these further two states (that is, the conscious representation of M^*) M^{**} . The fact that M^{**} is a conscious state entails, by this very analysis, that x has yet another pair of mental states — and we are off with the same old infinite regress. The fact that the analysis of Position 2' posits not *only* M_2 , but also another mental state, cannot curtail the infinite regress triggered by the fact that M_2 is conscious, and more generally, by the fact that the explanation of consciousness adverts to the very same phenomenon it purports to explain.³⁷

36 This difference shows up in the fact that the sub-argument against Position 1 only establishes an intermediate conclusion on the way to establishing Position 3, whereas the phenomenological argument is intended to establish Position 3 directly.

37 On top of this, the second reason offered in §V for rejecting Position 2 still applies: Position 2' is inconsistent with the unity of consciousness.

Sixth objection. The overall argument for Position 3, or (SA₃), is an argument by elimination. This mode of argument only works if all the options have been taken into account; otherwise the argument is invalid. And there is one option that the discussion in §V missed, namely, that in *some* cases the right model for consciousness is the one proposed in Position 3, but in other cases it may be the one proposed in Position 1 or 2.

This objection is correct in pointing out that this is a coherent option which the argument of the previous section has failed to take into account. At the same time, the fourth option here proposed is wildly implausible. In all probability, all intransitively self-conscious states share some feature, in virtue of exhibiting which they *are* intransitively self-conscious. It is improbable that there is nothing common to all intransitively self-conscious states, such that some are intransitively self-conscious in virtue of exhibiting feature F, some in virtue of exhibiting feature F*, etc. This response can be backed by the plausible claim that the class of intransitive self-conscious states forms a natural kind and must therefore exhibit some unity. So while the objector is right that the argument of the previous section missed a coherent option, the option in question is easily eliminated.

Seventh objection. The sub-argument against Position 1 depended on the premise that our awareness of our conscious states is something we are acquainted with in our first-person experience. But someone may defend Position 1 by insisting that she simply does not find this awareness in her experience.

My reply to this objection is threefold. First of all, defending Position 1 in this way may prove a slippery slope. The challenge facing the objector is to deny that we have first-person experience of an awareness of our conscious states *without* denying that this awareness of our conscious states is a real thing. For the objector is a defender of Position 1, and is therefore not an *eliminativist* about intransitive self-consciousness. Position 1 is, after all, a position on the nature of intransitive self-consciousness. The objector must therefore defend the existence of the relevant mode of self-awareness, but maintain that we have no first-person experience of it. As I said in the last section, it is relatively easy to envision someone who professes not to experience any form of implicit, non-reflective self-awareness and goes on to reject the existence of such self-awareness as a philosophical myth. It is much harder to imagine someone with similar phenomenological professions who insists on there being such self-awareness nonetheless. This latter character is our objector. The straightforward eliminativist position about intransitive self-consciousness is more plausible, but as I said at the end of §I, in this paper I do not set to disprove it.

Second, and relatedly, to defend Position 1 in this way, the objector would have to come up with certain psychological and/or behavioral data of which the best explanation entails positing second-order representations of our conscious states. To my knowledge, no one has yet pointed out any data whose explanation calls for positing such second-order representations.

Finally, recall that the literature is abundant with many other difficulties for Position 1, the position of the Higher-Order Monitoring theorist. Even though I think none of these capture the fundamental implausibility of the approach — the source of its disattraction, if you please — they certainly have a cumulative effect. So even if the main argument I have presented against it failed, there would be any number of other reasons to reject it.

VII Conclusion: Intransitive Self-Consciousness and Self-Representational States

In this paper I discussed the structure of consciousness in one particular sense of the term, namely, consciousness as intransitive self-consciousness. In this sense, a mental state is conscious just in case the subject has it self-consciously, that is, if she is implicitly aware of the occurrence of the state in question. This raises the question whether this awareness constitutes a distinct mental state, or is built into the very conscious state of which the subject is aware.

Brentano argues that this awareness is built into each conscious state itself, such that conscious states are always self-representing. The argument he offers proceeds by eliminating (i) the possibility that the subject's awareness of her conscious state is anchored in a non-conscious state, on the grounds that there are no such states, and (ii) the possibility that it is anchored in a separate conscious state, on the grounds that that would lead to infinite regress. This argument is unsound, however, because it is false that there are no non-conscious mental states.

I have offered an amendment to Brentano's argument, intended to dispense with the premise that there are no non-conscious mental states. The amended argument eliminates (i) the possibility that the subject's awareness of her conscious state is anchored in a non-conscious state, on the grounds that we as subjects have a first-person *experience* of this awareness, which therefore must be conscious, and (ii) the possibility that it is anchored in a second conscious state, on the grounds that it would lead to infinite regress (as well as on the grounds that it is inconsistent with the unity of consciousness). These possibilities being the only remotely plausible alternatives, it follows that the subject's

awareness of a conscious state of hers is anchored in that very state. Thus, when we self-consciously perceive the blue skies, our perception represents both the skies and itself.³⁸

Received: May, 2002

38 I would like to thank Gary Bartlett, Elizabeth Vlahos, and Ken Williford, as well as two anonymous referees for the *Canadian Journal of Philosophy*, for helpful comments on earlier drafts of this article.