David Bitter

MA THESIS

Department of Philosophy
Central European University

Budapest 2010

THE DISSOCIATION OF CONSCIOUSNESS

By
David Bitter

Submitted to
Central European University
Department of Philosophy

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Professor Howard Robinson

Budapest, Hungary
2010

Abstract

It is both clinically and experimentally established that hypnosis is highly effective in alleviating pain. Yet while hypnotic analgesia alters the subjective experience of pain, using certain special techniques, information registered on the pain outside of conscious awareness can be tapped by the hypnotist, eliciting "automatic reporting" on the pain by a "hidden part" of the subject – the "hidden observer".

Insofar as the hidden observer phenomenon involves two contradictory reports on pain (an "overt" report of analgesia and a "covert" report of pain), the possibility arises that the phenomenon constitutes a genuine empirical case of a dissociation of consciousness. Yet the idea of a divided consciousness seems both counter-intuitive and philosophically problematic. Thus, it is only expected that one should seek alternative explanations.

The central tenet of this thesis is that there is no easy way out of the apparent dilemma created by the hidden observer phenomenon. While the phenomenon doesn't seem to be explainable away as an experimental artifact, and conceptual-terminological issues of pain do not factor as relevant, all positive accounts of the phenomenon would seem to force us to let go of at least one of our "folk" intuitions concerning the nature of consciousness and/or pain.

The general conclusion of the thesis is that, ultimately, one is faced with a choice between three undesirable options: the first involves letting go of our conception that phenomenality is essential to pains and/or voluntary behavior (zombie model); the second involves letting go of our conception that our conscious experiences are (generally) temporally continuous (switching model); while the third involves letting go of our conception that our conscious experiences are under all circumstances unified (dissociation model). Each option has its advantages, but each comes with a bundle of problems as well – there is no easy way out.

Table of Contents

1. Introduction

1.1. The "hidden observer"

In a famous series of psychological studies conducted at the Stanford Laboratory for the Study of Hypnosis during the 1970s, Ernest R. Hilgard and his colleagues (Hilgard, 1973; Hilgard, Hilgard, Macdonald, Morgan and Johnson, 1978; Hilgard, Morgan and Macdonald, 1975; Knox, Morgan and Hilgard, 1974) provided evidence for a mind-boggling phenomenon. In a typical experiment, they would induce hypnotic analgesia in the study participants (i.e., a state or condition in which the subjects, despite their being awake and conscious, would feel no – or a significantly reduced level of – pain), after which they would immerse one arm of the subjects in ice cold water. As expected, most subjects reported no or very little pain upon questioning, and there were no visual or behavioral signs of discomfort. The participants were relaxed, their facial expressions manifested no hints of suffering and their behavior was totally consistent with their verbal reports. Yet, by means of certain special techniques, they succeeded in eliciting "covert" reports on "concealed information" about pain that was allegedly both registered and reported by the subjects outside of conscious awareness. Surprisingly, in spite of the subjects' firm allegations of not feeling any pain, the level and pattern of pain covertly reported matched both the external stimulus state of affairs and the subjective reports of normal wakeful subjects who were not hypnotized. Since subjects in these experiences provided no overt reports of pain, the concealed information and the method by which it could be tapped was dubbed the "hidden observer".

Strictly speaking, the hidden observer phenomenon (HOP) was not discovered by the Stanford research group; they merely rediscovered a phenomenon that was long before known to psychologists familiar with hypnosis and hysteria. For example, Alfred Binet, who devised the first ever IQ test, devoted a whole book (*On Double Consciousness*) to a

discussion of such phenomena already in 1890. William James published his seminal work, *The Principles of Psychology*, in the same year, in which he recounted an incident of a man whose hand was totally anesthetic while engaged in "automatic writing"[1] – yet while James could severely prick the anesthetized hand without the man exhibiting any awareness of this action, surprisingly, the writing accused James of hurting the hand (1890/1983, pp. 205-6). On the basis of such observations, James emphasized that

> during the times of anaesthesia, and coexisting with it, *sensibility to the anaesthetic parts is also there, in the form of a secondary consciousness* entirely cut off from the primary or normal one, but susceptible of being *tapped* and made to testify to its existence in various odd ways. (p. 201; italics in original)

Thus, he concluded that

> *in certain persons*, at least, *the total possible consciousness may be split into parts which coexist but mutually ignore each other.* (p. 204; italics in original)

What James meant exactly by "secondary consciousness" is an issue of considerable debate (cf. Weinberger, 2000), and as we shall later see, hidden observer (HO) experiments have provided evidence that even if the envisaged dissociation of consciousness is possible, the split off parts need not necessarily be mutually ignorant of each other.[2] Yet nevermind the details of all of this. My question is, rather, whether James' conclusion really follows from phenomena like the HOP. I am not convinced that it does. Then again, as I will argue, there seems to be no ideal alternative at hand – however we might go about explaining the phenomenon, we are bound to find trouble. Thus, the central tenet of this thesis is that, beyond challenging our central notions of pain, one way or another, the HOP might ultimately force us to give up some of our most cherished intuitions concerning the unity and/or temporal continuity of our conscious experiences.

[1] "Automatic" insofar as, *purportedly*, the subject was not consciously aware of his own writing.

[2] Strictly speaking, James didn't claim either that the split off parts of consciousness need necessarily mutually ignore each other, only that this *may* be the case.

1. 2. Various concepts of consciousness

If we are going to discuss the possible dissociation of consciousness, we should first specify what we mean by consciousness and by dissociation. Let us start with consciousness. Block (1995/1997) argued that the concept of consciousness is a "mongrel" concept insofar as it connotes multiple concepts and relates to various different phenomena. In line with Block, I will use the term *phenomenal consciousness* (P-consciousness) to refer to our phenomenal experiences, i.e., the phenomenal contents or the phenomenal aspects of conscious contents or states that render conscious states something "it is like" to be in those states (see also Nagel, 1974). The paradigm examples of P-conscious contents are sensations such as the redness of a rose or the sweetness of chocolate. Other paradigmatic examples are the qualitative properties of emotions such as the pleasurable character of happiness, or the phenomenal properties of hybrid emotion-sensations like pain. I assume that complex feeling-thinking states such as believing and understanding (might) have P-conscious contents/properties as well (cf. Harnad, 2001).[3]

Also in line with Block, the term *access-consciousness* (A-consciousness) will be used to refer to states the representational contents of which are (1) poised to be used as premises in reasoning (they are "inferentially promiscuous"), (2) poised for the rational control of behavior, and (3) poised for the rational control of speech (verbal report). Arguably, while the three conditions are jointly sufficient for A-consciousness, neither is individually necessary. As opposed to P-consciousness, A-consciousness is essentially a functional concept, closely related to the "global broadcast" / "global workspace" theories according to which (A-)consciousness consists in, or arises when, the representational outputs of various specialized cognitive systems (modules) are "broadcast" to a "global workspace" equipped

[3] I do not simply mean that thoughts are (might be) accompanied by, say, "inner speech", but that there is (or might be) something it is like to hold a belief or to grasp the meaning of something etc.

with a working memory, by virtue of which the representational contents become available as input to the various "consuming" systems (e.g., memory, planning, assessment, reporting or reasoning) (cf. Baars, 2002). It is important to notice that A-consciousness is a dispositional concept: a cognitive representation will be rendered A-conscious by virtue of its being globally available, which does not entail that it is actually "consumed" by any output system.

While there is a clear conceptual distinction between A-consciousness and P-consciousness, it is questionable whether it is empirically possible for the two to ever come apart. One intriguing aspect of the HOP is that, *prima facie*, it presents itself as a potential empirical candidate for just such a divide. According to one interpretation, subjects in the HO situation might actually be P-conscious of pain, yet overtly report not being so due to their incapacity to cognitively access this pain (see *Section 4.1.*). According to another interpretation, perhaps subjects really are analgesic, and their capacity to provide accurate covert reports on the stimulus state of affairs constitutes a genuine example of A-consciousness without P-consciousness (see *Section 4.2.*).

Of course, there is a sense in which P-conscious states/properties must be accessible. Consider pain, for example: would it make sense to claim that there is something it is like for a pain to be if there were no subject *for whom* it was like something to experience the pain? Arguably, no, it wouldn't. So I will assume that if a state or experience is P-conscious, there must necessarily be a subject for whom it is like something to be in, or to undergo, that state or experience – or as James eloquently put it, "The universal conscious fact is not 'feelings and thoughts exist', but 'I think' and 'I feel'" (1890/1983, p. 221).

Note that adhering to the above claim in no way commits one to any particular metaphysical position on the nature of the subject (Levine, 2007). Neither does the claim entail that for a subject to be conscious, she must also be self-conscious or meta-conscious. Take for example

"flow" experiences (Csíkszentmihályi, 1990). Assumedly, people undergoing such experiences are P-conscious. Yet it is an essential characteristic of flow that while one is fully absorbed in a particular activity, she loses her sense of self-consciousness – if one were to focus on her own state/experience, this would inhibit/terminate the very experience of flow. Of course, since it is also a characteristic of flow experiences that one is engaged in the successful execution of some task/activity, such states also involve A-consciousness. So, assumedly, one might be both A-conscious and P-conscious, yet cease to be self-conscious.

1.3. Various concepts of dissociation

There are many ways in which the term dissociation might be understood. For example, cognitive neuropsychologists often speak of dissociations in the sense of selective impairments in mental functioning. Such impairments are selective insofar as there are certain particular tasks or domains in which one function might be impaired, yet another spared. A well-known example is that of blindsight: while one visual system (the striate cortex) of patients afflicted with this neuropsychological syndrome is damaged, thus preventing conscious access to visual stimuli, another visual system (the superior colliculus) is spared, thus enabling patients to provide above-chance guesses concerning, e.g., the identity of objects in their blind field (Weiskrantz, 1986). Another interesting example is the dissociation between tonal pitch perception and general pitch perception: while in the case of auditory atonalia, recognition of melodies is ruptured yet perception of pitch distances and directions is spared, the reverse is true for generalized auditory agnosia (Peretz and Hyde, 2003). Interestingly, pain itself seems to be comprised of multiple components which, under certain conditions (like hypnotic analgesia), might also dissociate in this sense of the term. Yet, as I will discuss the issue in *Section 3.*, this type of dissociation will hardly provide a full and satisfactory account of the HOP.

Further notions of dissociation relate not to impairments, but to the disunity of A-conscious contents and/or the disconnectedness of various intact functions. Following Bayne and Chalmers (2003), we might say that two conscious states are *access-unified* when they are jointly accessible. In this sense, if the distinct contents (say, A and B) of two distinct mental states are both individually accessible, yet there is no such representation accessible to the consuming systems with the content *that* A&B, then these states will be access-*dis*unified. The neuropsychological syndromes discussed above do not involve dissociations in this sense, for in those cases the impaired function is not even individually accessible. On the other hand, disconnection syndromes like schizophrenia seem to be good examples of such access-disunity insofar as persons suffering from schizophrenia might have individual access to various perceptions, thoughts, motives etc., yet be incapable of integrating (jointly accessing) these contents for the purposes of planning, initiating and executing organized behavior. Yet the HO is in no way like persons suffering from schizophrenia: she manifests consistent and organized behavior and seems to have meta-access both to her own states/experiences, and the states/experiences of the "hypnotized part" (HYP) of the subject (viz., the "part" which provides the overt reports).

Thus, in important aspects, HOs seem neither impaired, nor disconnected in the sense of disintegration/non-integration. Yet there are well documented cases of disconnection in which A-consciousness seems to be split into two or more (at least somewhat) distinct parts, with each part itself manifesting unity and integration. Take for example persons suffering from "word meaning deafness". Such persons, while unable to understand spoken words, are nevertheless both capable of repeating spoken words *and* of accessing semantic information in the visual modality. This suggests that both the acoustic-phonological system dedicated to analyzing spoken word and the system devoted to semantic analysis are (at least partially) spared in this condition (Ellis and Young, 1988, *in* Schacter, 2000). Note that word meaning

deafness is not merely a disconnection in cognitive functioning at the sub-personal level; persons suffering from this condition are individually A-conscious of both the acoustic-phonological properties of spoken words and their semantic meaning (in the visual modality). Thus, in such cases, it seems that A-consciousness itself has split into two separate parts or systems, each of which is internally unified and well-functioning.

Conceptually, there is nothing especially controversial about the notion that A-consciousness might split into two (or more) distinct yet internally unified parts; it is perfectly coherent to assume that while A and B are both individually and jointly accessible to one set of consuming systems, C and D are both individually and jointly accessible to a different set of consuming systems. Indeed, such disconnectedness is exactly what is so commonly observed in "split-brain" patients. In such patients, the cerebral commissures connecting the two hemispheres of the brain have been sectioned, most often for the purposes of confining epileptic seizures to one side of the brain, thus alleviating the epileptic symptoms. While such patients generally exhibit ordinary behavior in everyday life, numerous distinct impairments due to a lack of integration have been demonstrated under special experimental conditions. For example, in a typical experiment, the word "catkin" is visually projected on a screen in such a way as to ensure that the right hemisphere of the patients' brain (which receives input from the left visual field) will only receive the visual input "cat", while the left hemisphere (which receives input from the right visual field) will only receive the visual input "kin". The intriguing finding is the following: while patients can only verbally report the word "kin" (for their speech centers are localized to the left hemisphere), and further, while they firmly assert not having seen or being aware of any other word, with their left hand (which is under the control of the right hemisphere), they can nevertheless draw or pick out (say) a small toy cat. Yet they do not (and cannot) draw or pick out a catkin, for they are incapable of integrating

the two distinct representations ("cat" and "kin") into a third representation (Sperry, Gazzaniga and Bogen, 1969).[4]

It seems more than natural that the empirical evidence on split-brain patients caught the attention of philosophers. For if A-consciousness can split into two separate, fully functional systems in cases like the split-brain syndrome, then perhaps P-consciousness might mirror the cognitive divide and possibly split into two (or more) parallel streams as well? The question seems especially relevant for the philosophy of mind, for it seems that whichever way we attempt to answer this question, it will be impossible not to have to let go of at least one (or more) characteristic of consciousness that we generally regard as essential to it (see *Sections 1.4.* and *3.*). Since this issue is of great relevance regarding the HOP as well, I will now summarize the possible positions I believe one might take in relation to the issue of dissociated P-consciousness in split-brain patients.

One might assume a "Descartian" position and assign P-consciousness only to the left hemisphere (which possesses the linguistic capacities underlying verbal reportability), alleging that the right hemisphere is an "automaton". Yet the right hemisphere of split-brain patients seems all but: it can semantically grasp verbal instructions, it can perform complex cognitive tasks, and it can initiate what seems like voluntary behavior (Nagel, 1971; Sperry et al., 1969). In short, the right hemisphere seems A-conscious. Yet if this is so, and one were still to deny it P-consciousness, then the right hemisphere would seem to qualify as an empirical case of a "zombie".

[4] One might argue that the left hand could not have possibly drawn a catkin because, in order to combine "cat" and "kin", it would have had to first form a representation of "kin" – an apparently not easily visualizable/drawable concept. Yet this line of reasoning is ruled out by literally hundreds of other studies in which the visual stimuli projected to both hemispheres were easily visualizable/drawable.

Another option is to assume that, though both hemispheres can be both A- and P-conscious, neither of them are conscious at the same time (Bayne, 2008). According to this version, there is one attention system which switches its attention back and forth between the two hemispheres. The strength of this view is that it seems to save our intuition concerning the empirical impossibility of zombies without positing the dissociation of consciousness. Then again, it is highly counter-intuitive in the sense of pitting itself against our subjective impression that, generally, our conscious experiences are temporally continuous.

If one were to be unsatisfied with the automaton theory, the zombie account and the switching model, one might still opt for Bayne and Chalmers' (2003) notion of subsumption, claiming that, for any set of P-conscious mental states, the P-states associated with such mental states will necessarily be subsumed by a single P-conscious state. Notwithstanding the issue of by virtue of what such phenomenal states would or should be subsumed under one total unified P-state in cases of dissociated A-conscious states, as I see it, there are two further apparent problems with this account. First, in cases of competitive behavior and/or cognition, it will go against our intuition that P-unified experiences within a single modality must be representationally consistent (Bayne, 2007). Second, it is at least highly controversial whether such a position really doesn't entail the dissociation of consciousness. Bayne and Chalmers (2003) admit that "there is a sense in which a breakdown of access unity is a 'disunity' in consciousness," yet they believe this is so only "in a relatively shallow sense" (p. 17). Of course, "shallowness" is a relative term; nevertheless, in *Section 4.1*., I will argue that, at least in the case of the HOP, the subsumtive thesis entails dissociation in quite a "deep" sense.

It might also be argued that insofar as the subsumptive thesis posits the existence of never-in-principle-accessible P-conscious states in the case of A-dissociation, such P-states would be

more akin to "free-floating" P-states or "isolated conscious mental phenomena, not integrated into a mind at all, though they can perhaps be ascribed to the organism" (Nagel, 1971, p. 403). Yet, as I have noted in *Section 1.2.*, the subjective aspect of P-consciousness seems essential to it: "phenomenal states/properties are not merely instantiated *in* the subject, but are experienced *by* the subject" (Levine, 2007, p. 514; italics added).

If I am correct, there are three positions left, all of which assume that the split-brain person has two distinct P-conscious states which, while they might have subsumed other P-states, they themselves are not P-unified with each other. Those who agree that P-consciousness entails a subject of experience yet see something essentially incoherent or inconceivable in the notion that a single subject might have dissociated P-states, will posit that if there are two non-unified P-states, then, necessarily, there will be two subjects as well. On the other side, those for whom the idea of a single subject splitting into two is inconceivable will have to posit that there is a single subject of experience(s) who is independently P-conscious of both A and B, yet for whom there is nothing it is like to jointly experience A&B. While strong versions of such a position might, e.g., construe of an immaterial self whose consciousness could be split into two without the self noticing this (see, e.g., Robinson, 1989), milder versions of this view ("partial unity" theories) will posit, beyond A and B, a third P-state C, with which both A and B are co-conscious without themselves being unified (Lockwood, 1989, *in* Bayne, 2009; cf. also Brook and Raymont, 2009).[5]

[5] I have not included in my summary Puccetti's (1981) position on dual consciousness, for insofar as he believes that even non-split-brain persons have two parallel streams of consciousness, his view implies a dissociation only in the accessible contents and not the structure of P-consciousness. Neither have I presented views according to which split-brain persons generally have one consciousness which only dissociates in certain special circumstances (cf. Bayne, 2008; Nagel, 1971), for my interest lies not in the relative frequency, but the very possibility of a genuine dissociation.

I have only briefly mentioned the various views of dissociation and commented more on those positions which deny that we need to posit that split-brain patients have two co-existing streams of P-consciousness because the primary question of this thesis is whether it is possible, in the first place, that the HOP might constitute a genuine case of dissociated consciousness. Thus, by the term "dissociation of consciousness", I will not distinguish between views that attribute multiple streams of consciousness to a single self from views which adhere to a "one subject – one unified consciousness" view; if I were to conclude that either version is possible, I would equally assume the possibility of genuine P-dissociation. As for the potential coming apart of A-consciousness and P-consciousness, though I do not take it to be an evident matter, as I will argue that "phenomenal overflow" arguments entail P-dissociation, potential cases of P-consciousness without A-consciousness will be considered cases of conscious dissociation as well.

1.4. The relevance of the hidden observer phenomenon

It is commonly held that the most convincing evidence for a possible dissociation of consciousness comes from split-brain studies (Bayne, 2007; Nagel, 1971). Yet there are multiple factors that constrain the interpretation of split-brain data. First, since the right-hemisphere is barred access from the areas underlying language production, it is impossible to attain verbal reports from split-brain patients. A clear advantage, then, of the HOP is that HOs can produce both verbal and written reports concerning the states they are undergoing. Second, since the split-brain phenomenon arises from structural/neural damage in the patients' brain, the condition is not reversible: patients will never be able to report even retrospectively about their experiences in relation to their right hemisphere. Yet, while the credibility of retrospective reports concerning P-conscious experience is at least questionable,

the temporary (reversible) nature of hypnotic phenomena do provide further important data on the kinds of psychological processes split-brain patients might be undergoing.

Some believe that the fact that hypnotic subjects have undergone no operation, i.e., that their brains are neurologically intact, weakens the case for a dissociation of consciousness in hypnosis, for "it is unclear what feature of the hypnotic context in general – or hidden-observer paradigms in particular – might change the structure of consciousness" (Bayne, 2007, p. 99). Yet this kind of reasoning seems to confound the issue of descriptive validity (i.e., whether there really is a dissociation taking place in the HOP) with the issue of explanation. Further, it is not clear why physical/structural damage of the brain should be conceived of as so essential to P-dissociation in the first place. Also take note: just because dissociative phenomena under hypnosis are "functional" in nature, this in no way implies that there are no underlying neural changes associated with such conditions (Kihlstrom, 2005).

Then again, one might accept the above points and admit that "even if consciousness in the split-brain syndrome remains unified, it is possible that the unity of consciousness breaks down in the context of other pathologies of consciousness; and, of course, it is possible that the unity of consciousness might fail in non-human animals" (Bayne, 2008), yet nevertheless cling on to the claim that, since hypnosis is not a pathological condition, there is no reason to believe that it involves a dissociation of consciousness. Such an argument would not seem to have much philosophical import. If one were to adhere to the view that the dissociation of consciousness in humans is necessarily associated with pathology, then so be it: perhaps hypnotic dissociations (if possible) are "pathological". While psychologists would tend to disagree with such a claim – arguing that *even if* the existence of hypnotic states could or should be construed of as "the basis and determination of hysteria" (Breuer and Freud, 1895/2007, p. 8; cf. Spiegel, 1990), this in no way implies that hypnotic states *per se*

constitute a breakdown of normal psychological functioning –, such disagreement about "pathology" would rather constitute a terminological than a substantive debate about the nature of hypnosis.

In split-brain studies, the two hemispheres of the brains of the subjects are typically confronted with different physical stimuli: e.g., the visual projection of the word "cat" is "presented" to one hemisphere, while the word "kin" is presented to the other. A further advantage of the HO paradigm, then, is that subjects are exposed to the same physical state of affairs, with the HO and the HYP asked to report on the same thing – the level of pain experienced. This feature of the HOP brings out (and challenges) our intuition concerning the necessity of intramodal consistency of phenomenal and/or representational content/properties in a way more direct manner than any evidence so far from split-brain patients. For even if one were ready to make concessions on the issue of consistency based on split-brain cases, there seems to be something essentially inconceivable about the notion that one could, at any particular given time, both be in pain and not be in pain.

Finally, not only are the reports of the HO and the HYP in direct conflict within the HO situation, but as opposed to split-brain cases, where the two hemispheres of the brain are mutually ignorant of each other, the HO seems to have knowledge of the internal goings on of the HYP as well. (Note that the relation of meta-access is asymmetrical: while the HO is "omniscient", the HYP is ignorant of the HO).

Thus, not only does the HOP point towards the possibility of the dissociation of consciousness, but it seems to provide more kinds of – and perhaps even more baffling – data than split-brain patients on the possible internal goings on of persons exhibiting cognitive (A-conscious) dissociations. In view of this, it is interesting that while split-brain phenomena have received extensive treatment in the philosophical literature, there is hardly any reference

to – not to mention proper discussions of – the HOP in philosophical works (exceptions are Bayne, 2007; Block, 1995/1997).

The HOP also raises important issues about our concept of pain. As noted above, it seems hardly conceivable that a single subject could have a unified P-experience of both being in pain and not being in pain (*consistency claim*). Yet if (i) the overt reports of HYPs (concerning no pain) are credible (*validity claim*; see *Section 2.1.*) and (ii) the covert reports of HOs cannot be explained away as mere psychological automatisms (*controlled behavior claim*; see *Sections 2.2.* and *4.2.*), yet (iii) one were inclined to deny the empirical possibility of P-dissociation (*P-unity claim*; see *Sections 3.* and *4.1.*) and nevertheless (iv) cling on to the *consistency claim*, then it seems that (v) one would either have to (a) posit the empirical possibility of "partial zombies" and thus let go of the conception that pains are necessarily "painful" (*phenomenality claim*; see *Sections 3.* and *4.2.*), or else (b) let go of the conception that we have access, under all conditions, to our own phenomenal pains (*accessibility claim*; see *Sections 3.* and *4.1.*), in effect implying that we should also let go of our intuition of *incorrigibility* (see *Section 3.*).

As I will argue in *Section 5.*, the only way to break out of this apparent dilemma is to assume that, even though the P-experiences of both the HYP and the HO seem to be *continuous* to the subjects, from a metaphysical point of view, they are not. Yet, since this view entails that subjects in the HO situation are victims of a "hyper-illusion"[6] concerning the uninterrupted flow of their P-consciousness, not only would such a position force us to abandon our notion of the temporal continuity of consciousness, but it might also have dire consequences concerning our notion of subjects persisting through time.

[6] Block (2007) calls "hyper-illusion" the kind of illusion which involves not a mismatch between an appearance and the external world, but a mismatch between an appearance and an appearance of an appearance. In his words, a hyper-illusion is a "kind of illusion in which the introspective phenomenology does not reflect the phenomenology of the state being introspected" (p. 493).

Finally, it should be noted that the issue of whether the HO feels any pain is not merely an issue of theoretical relevance for the philosophy of psychology or the metaphysics of mind. For if the HO really feels pain while the HYP does not, then this would seem to have significant ethical implications as well, especially for the clinical application of hypnotic analgesia.

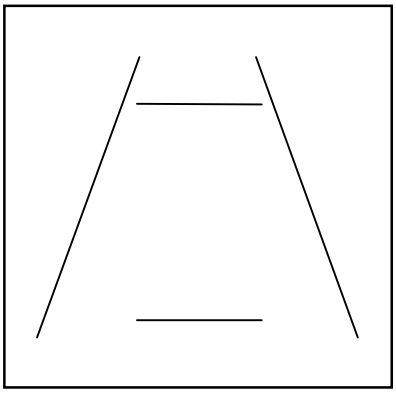
2. The validity of the hidden observer phenomenon

2.1. Hypnotic analgesia

While demonstrations of what are called *negative hallucinations* in psychology are most dramatic in the case of hypnotic analgesia, as noted in the introduction, the phenomenon is much more general. For example, modern research has also provided evidence for hypnotic deafness (Hilgard, 1977) and hypnotic blindness (Zamansky & Bartis, 1985). Shall we give credit to such evidence? I believe we should – yet this will need a fair amount of arguing.

Prima facie, there would seem to be plenty of reason to believe that in the HOP, subjects have, after all, perceived the pain. First of all, it is well established that negative hallucinations do not eliminate sensory detection of the presented stimuli. For example, James (1890/1983, p. 607) mentions that if we ask a subject, for whom a red cross on a sheet of paper is apparently invisible, to fixate his eyes for a while on a particular point on the paper, then, upon looking at a blank paper, the subject will see a bluish-green afterimage of the cross. Further, take the example of the Ponzo illusion (see *Figure 1*.). Under hypnosis, it is allegedly possible to selectively ablate perception of the converging lines, yet not of the horizontal lines. Yet the illusion persists, and subjects still report seeing the upper line as longer than the bottom line. This raises the so called "paradox of the Judas eye" (Hilgard, 1977; Kihlstrom, Barnhardt and Tataryn, 1992): for in order to selectively ablate the background in the picture, one must first identify and distinguish (i.e., perceive) it. Perhaps not surprisingly, then, HOs have been reported to be covertly perceiving the converging lines (Gettinger, 1974, *in* Hilgard, 1977).

Figure 1.



The Ponzo Illusion

No doubt about it, it seems logically contradictory that one might both be seeing/feeling something yet not be seeing/feeling that something. An understandable skeptical reaction, then, is to assume that the subjects in these experiments are in reality perceiving everything – perhaps HYPs experience pain all along, and they only *claim* not to be suffering. After all, with a bit of practice, one might easily fool the hypnotist. So perhaps there is nothing "hidden" about the HO and her reports represent the actual experiential state of affairs.

Yet why would experimental subjects lie? Why would they claim not to be feeling pain if, in fact, they were suffering all along? One possible answer is that they were merely complying with the "demand characteristics" (Orne, 1959) of the experiment. The expression refers to the unstated hypotheses that the experimental procedure might implicitly convey. The argument, then, is that since subjects are assumedly motivated to respond positively to the suggestions of the hypnotist, it is rather possible that they engage in strategic self-presentation in order to please the hypnotist as a "good" subject (Spanos, 1986; Wagstaff, 1981, *in* Kallio and Revonsuo, 2003).

Indeed, so called "simulators" – subjects who are not susceptible to hypnosis, yet who are instructed by another experimenter to behave during a hypnotic session in a way they think is consistent with how the hypnotist expects them to behave –, can easily fool the hypnotist in the HO experiments. Some simulators might even claim not to have felt any pain during

interviews taken *after* the hypnotic session. Then again, when subjects are prompted to give honest reports and/or questioned in a different context (and not by the hypnotist), simulators *do* admit having felt pain all along, while "reals" uphold their claim of having felt no pain (Hilgard et al., 1978).

The skeptic might object in the following way: since (i) the simulators can clearly pick up on the demand characteristics of the post-hypnotic interview to produce honest reports, and since (ii) they are well aware of the fact that at least one of the experimenters (viz., the one who selected and instructed them to simulate) knows about their identity, and further, since, (iii) it is only understandable that, ultimately, they would want to avoid being seen as immoral agents, it is only to be expected that, in the "honesty interviews", simulators would (yet again) comply with the situational demands for honesty. On the other side, insofar as (i) reals have no "cover story" for why they might have faked their analgesia, and insofar as (ii) they can be quite sure that nobody (besides them) knows about their "faking", since (iii) it is only understandable that they would want to avoid seeming inconsistent, it is only to be expected that they would keep up with their "narrative" of the analgesic subject. (for a similar argument, see Spanos and Hewitt, 1980).

Not only is the above line of reasoning absurd, but there is also empirical evidence against it. I claim that the reasoning is absurd because it implies that it is the reals who are really the deceivers. Think about it: simulators are selected for their task on the very basis of having "admitted" after previous hypnotic sessions that they didn't undergo the experiences "expected" of them. Yet if one assumes that reals didn't really undergo such experiences, either, then, apparently, what distinguishes between reals and simulators is that the former fake their experiences more persistently and more convincingly.

So here's two pieces of empirical evidence against the claim. First, using a single measure (electrodermal skin conductance) for detecting deception, a team of researchers (Kinnunen, Zamansky, and Block, 1994) found that, when asked for honest reporting, 89% of reals met the pre-set criterion for truthfulness, while only 35% of the simulators' reports met that criterion. True, sociopaths and undercover agents can fool a lie detector as well. Yet high-hypnotizables are hardly sociopaths, and it takes several months of intensive training (even for talented deceivers) to acquire the skill of deceiving a lie detector. Second, in their honesty reports, reals readily and commonly report on occasional intrusions of pain, while simulators typically overact their role, providing lower overall ratings of pain and no reports of intrusions (Hilgard, 1977). Arguably, it is hardly the case that reals intentionally manipulate their reports in such a way as to "avoid" being "caught" on "cheating".

If someone were still not convinced, consider the fact that hypnotic analgesia is successfully applied, among others, in obstetrics, dentistry, the treatment of postoperative pain, chronic pain, and pains associated with burns, cancer and cancer treatment (Hilgard and Hilgard, 1994; Montgomery, Duhamel and Redd, 2000). Shall we assume that persons in such situations are "faking" and merely complying with the "demands"? What could their motivation be for doing so? Here's what Hilgard (1973, p. 404) had to say about this:

> On both the phenomenal and social side, it is only necessary to call attention to the woman who having had one child with the help of hypnotic analgesia requests that she have hypnosis again for her next child (even though the original obstetrician is not around), or to one of our own subjects for whom the local anesthetic used by the dentist has not relieved pain sufficiently but now gains complete pain reduction in the dental chair through self-induced analgesia without any chemicals injected by the dentist. No hypnotist is around either to embarrass or to please.

Thus, while situational and social factors, as well as the motivations, attitudes and expectations of the subjects can and do influence the effects of hypnotic analgesia, this in no way implies that there is no genuine pain-relief in such cases (Kihlstrom, 1998; Kihlstrom

and McConkey, 1990). On the contrary, in a study comparing various methods of pain relief, hypnosis was found to be more effective than acupuncture, placebo acupuncture, morphine, aspirin, diazepam, and placebo in counteracting both cold-pressor and ischemic pain (Stern, Brown, Ulett and Sletten, 1977, *in* Kihlstrom, Mulvaney, Tobias and Tobis, 2000). Interestingly, at least in persons highly susceptible to hypnosis (Kallio and Revonsuo, 2003), this effect is not mediated by relaxation (Faymonville, Roediger et al., 2003; Miller, Barabasz and Barabasz, 1991), endogenous opiates (Goldstein and Hilgard, 1975), placebo effects (Hilgard and Hilgard, 1994), stress-inoculation techniques like diverting of attention (Miller and Bowers, 1993) or counterpain imagery (Hargadon, Bowers and Woody, 1995). Of course, the question of how, then, this effect is achieved, is not relevant here. The important point is that hypnotic analgesia *can* relieve pain, and that at least one component of this effect seems to be genuinely attributable to hypnosis.[7]

2.2. The "hidden observer"

The evidence and arguments in favor of hypnotic analgesia is so overwhelming that the majority of role-playing theorists have by now long abandoned their original positions on faking, conceding that subjects' reports of pain relief generally reflect genuine experiences.[8] It only seems natural, then, that the skeptic who would wish to undermine the validity of the HOP will now turn to attacking the credibility of the HO's reports.

[7] I have not yet defined hypnosis. Though there is no universally accepted definition, roughly, we might characterize hypnosis as a procedure taking place in a social interaction in which one person (the hypnotist) offers suggestions to another person (the subject) for experiences involving alterations in sensation, perception, emotion, thought or behavior. "Suggestions" are verbal or non-verbal communications that the hypnotist uses to guide the subject in undergoing the changes in experience (Kihlstrom, 1987; Kihlstrom and Barnier, 2005; Kirsch and Lynn, 1998). Yet note that the hypnotist can be the very person hypnotized: one can be taught to guide himself through a hypnotic procedure (Hilgard, 1977).

[8] Instead of denying the phenomena, revised versions of role-playing theories now posit that high-hypnotizables enact the social role of a hypnotically anesthetized subject by actively employing certain cognitive strategies in a way that, eventually, they may come to deceive themselves that they feel no pain (cf. Kihlstrom and McConkey, 1990; Kallio and Revonsuo, 2003).

A question that arises instantly is the following: if it was already conceded that subjects generally provide honest reports in the case analgesia, then why assume selectively that perhaps the HO is not telling the truth? The role-theorist might argue that insofar as the subjects really did achieve genuine pain relief, there was no room or need for faking up till the point at which the HO was called forth. Yet, at this point, either the HYP would have had to "click out" of her analgesic state, or else she couldn't have felt any pain. In light of the evidence on hypnotic analgesia, the second option would seem like a safer bet. So, since the skeptic is still assuming that subjects have a strong motivation to please the hypnotist, it would seem to follow that the HO might have falsely reported on pain just in order to "please" the hypnotist.

A well-known and much debated study by Spanos and Hewitt (1980) might seem to lend some support for the above interpretation. In this study, the subjects were divided into two groups: one group received the traditional suggestion that they would have a "hidden part" that was "more aware" of things happening in and around them, while the other group received the opposite suggestion, viz., that their "hidden parts" would be "less aware". Since, as the researchers expected, both groups gave reports that were congruent with the suggested level of awareness, they concluded that the "'hidden' reports result from Ss' [subjects'] attempts to convincingly enact the role of good hypnotic subjects" (p. 1201). Bayne (2007) agrees with this interpretation, arguing that the study underscores the point that *even* if the "more aware" HOs' reports were veridical, this might be accounted for by reference to the implicit cues contained in the experimental context, in effect challenging the claim that the HO reports on the actual stimulus state of affairs.

There are various possible objections to the above line of reasoning. First, while it is known that HOs are themselves susceptible to hypnosis (cf. Spiegel, 1990; Kihlstrom and Barnier,

2005), this neither excludes the possibility that the HO might, at least in principle, have access to otherwise inaccessible information "registered" by her cognitive system, nor the possibility that, even if the HO were merely an experimental creation, the HOP might still involve a genuine dissociation of consciousness (for veridicality is not the issue – simultaneous experiences are).

Second, since the wording and interpretation of suggestions play an important role in the way suggestions exert their influence, it is not at all clear why one should assume that the subjects responded in the way they did in the Spanos and Hewitt study *because* of merely complying with the experimental demands. Perhaps the suggestion of a "less aware" hidden part did lead to a dissociation of consciousness, yet, compared to the classic HOP, in this condition, instead of being omniscient, the divided part was even more anesthetized than the HYP.

Third, perhaps the very notion of a "less aware" hidden part made no sense to the subjects, and they really did provide the reports they did in order to conform to social pressure. This possibility brings to mind the famous line judgment study of Asch (1951), in which subjects in groups were presented with three (comparison) lines of differing length, and their task was to judge which of these matched in length with a fourth (target) line. While the answer was evident in each case, the setup was that in each group, there was actually only one real subject; the others were all confederates of the experimenter – a fact of which the participant was naïve about. Each subject had to give their answer aloud, in front of the whole group, and the naïve subject was always the last one. The experimental manipulation was that in two-thirds of the trials, the built-in "participants" gave consistent yet evidently wrong answers. The result: about one third of the participants conformed to the group pressure. The general moral of the study – at least according to Asch – is that people are prone to conform to social pressure *even when* they clearly know this will deter them from the correct answer.

The analogy with the HOP is clear: a subject's giving in to pressure or complying with demands does not entail that the subject had no access to the actual stimulus state of affairs (Kihlstrom and Barnier, 2005). Of course, while participants in the Asch studies acknowledged retrospectively that they had merely given in to group pressure, subjects in the Spanos and Hewitt (1980) study reported no such conformity. So, instead of pushing this analogy further, let us provide additional arguments and evidence in favor of the view that the HO's reports are credible and genuine.

First, it is important to note that only about 40% of highly hypnotizable persons (who in turn constitute 10-15% of the population) produce a HO (Hilgard & Hilgard, 1994). When both reals and simulators are put into high-demand situations, the incidence of HOs raises to about 50% among the reals and to 75% among simulators (Hilgard et al., 1978). Yet a study found that, when put into a low-demand situation, the percentage of real HOs fell back to the average rate, while the proportion of simulators" dropped down to zero (Nogrady, McConkey, Laurence and Perry, 1983, *in* Kihlstrom and Barnier, 1985). Thus, while reals *are* sensitive to demand characteristics, simulators are *way more* sensitive.

The skeptic might now say, fine, there are perhaps clear and detectable differences between reals and simulators – but this does not establish the validity or credibility of the reports of reals. For example, subjects might simply be confabulating (i.e., producing reports based on imagined/constructed false memories). Take the example of post-hypnotic suggestions combined with post-hypnotic amnesia. In such cases, the hypnotist might suggest to the subject under hypnosis that the subject perform a certain act (say, open the window) upon a prearranged cue (say, a cough) after the hypnotic session has terminated. The hypnotist will also suggest to the subject that, after the termination of the hypnotic session, he become totally amnesic to the very suggestion until another prearranged cue (say, two short coughs in

quick succession) is given, after which the subject will be able to easily recall everything. Sure enough, subjects for whom the first suggestion had an effect will open the window upon hearing a cough. When asked why they opened the window, they will be incapable of reporting the real cause of their action (viz., the suggestion); typically, they will confabulate something (i.e., produce a rationally sounding account based on an unintentionally constructed false memory). Of course, after the two short quick coughs, the subject will regain access to her memories, realizing the confabulated nature of her previously given answer(s).

Are the subjects in the HO paradigm also confabulating? This possibility is especially interesting, because Hilgard (1977) himself posited that the HO was inaccessible to the HYP due to an "amnesic-like barrier". The analogy arises from the fact that, both in the HOP and in post-hypnotic amnesia, it is clear that the subject possesses information (stored in her memory) which is not necessarily accessible (retrievable) at any given time. Yet the analogy seems to stop here. For, first, as opposed to cases of post-hypnotic amnesia in which a subject is barred access to previously had P-conscious experiences, Hilgard hypothesized that in the case of the HOP, certain states, experiences and/or information might be barred from becoming (P-)conscious altogether (i.e., information supposedly bypasses consciousness and gets written into memory directly.) Second, to return to the confabulation issue, amnesic subjects confabulate upon attempting to provide rational/coherent accounts of their own (observed) behavior. Yet the HYP has *no* access whatsoever to the doings of the HO; thus, there is nothing observed to rationalize or confabulate about. Third, in post-hypnotic amnesia, people confabulate *before* the amnesia is lifted. Yet in the HOP, subjects do not report anything on the HO until *after* their amnesia has been lifted – and once their amnesia is lifted, they are well-capable of giving veridical reports of the stimulus state of affairs they had been exposed to.

What is the basis of our claim of veridicality? Well, HOs' reports on the level of pain match both the stimulus state of affairs *and* the subjective reports of normal waking subjects who are not under hypnosis.[9] Yet perhaps subjects might have provided their veridical reports based on prior knowledge/acquaintance of the painful stimuli? Since, in the original HO studies (Hilgard, 1973; Hilgard et al., 1975; Hilgard et al., 1978; Knox et al., 1974), subjects had already been exposed to the painful stimuli in sessions preceding the session involving hypnotic analgesia, the possibility arises that the HOs were merely inferring the level of pain (which they *would* have felt had they not been in an analgesic condition) based on their prior experiences. Yet this worry is cleared by studies in which subjects had no practice sessions or prior exposure to the stimuli before the hypnotic session (Zamansky and Bartis, 1985). Yet perhaps subjects were not really reporting on pain, but on other tactile sensations (e.g., cold temperature) associated with pain? This isn't probable in face of studies which involved the induction of ischemic pain or electrical stimulation. Yet perhaps subjects might have inferred the level of pain due to prior exposure to the (kind of) stimuli from outside of the laboratory? Hardly so: for even if subjects were exposed to differing patterns of stimulation, HOs could provide accurate retrospective accounts of the location, duration, intensity and quality of the pain (Hilgard, 1977).

Hence, the HOP does not seem to be explainable away by appealing to role-enactment, demand characteristics, confabulation and/or prior knowledge or acquaintance of the stimuli. If the skeptic is to provide a satisfactory account of the phenomenon, then, it really seems like he will at least have to assume the face validity of the HOP.

[9] Actually, HOs report a bit lower level of pain than average waking subjects do – their ratings match those of subjects in a relaxed state (Hilgard, 1977; Hilgard and Hilgard, 1994). Yet this poses no problem insofar as being in a relaxed state modulates the very way in which sensory input from noxious stimuli is processed, and/or insofar as the very lack of the expression of pain via channels that are normally under voluntary control (e.g., facial expression) result in a lower level of experienced pain due to a lack of facilitative feedback from such expressive channels, the HOs' reports seem to match the level of pain they would feel in a similarly relaxed yet non-hypnotic state.)

3. On pain

In the previous section, I argued that we cannot explain away the reports of neither the HYP, nor of the HO, by assuming that subjects are merely faking or confabulating or complying with experimental demands etc. More specifically, we might summarize the purported conclusions of the previous section in the following way:

(i) If, within the hidden observer situation, a highly hypnotizable subject under hypnotic analgesia overtly reports that she is not in pain, then we have no reason to doubt that the subject, or at least the HYP, is *not* in pain.

(ii) If, within the hidden observer situation, a highly hypnotizable subject under hypnotic analgesia covertly reports that she is in pain, then we have no reason to doubt that the subject, or at least the HO, *is* in pain.

In order to bring out the deeply puzzling nature of the HOP, let us assume that if we have *no reason* to doubt the credibility of a subject's (or of a part of a subject's) report, then what is being reported matches the actual state of affairs. Based on the above, then, we might formulate the following assumptions (naturally, still to be understood as relating to highly hypnotizable subjects within the HO situation):

(i') If the HYP reports that she is not in pain, then the HYP really *isn't* in pain.

(ii') If the HO reports that she is in pain, then the HO really *is* in pain.

The puzzling fact is that, as mentioned earlier, in around 40% of the cases, the following are true:

(iii) The HYP reports that she is not in pain.

(iv) The HO reports that she is in pain.

Based on the above, then, we can conclude the following:

(v) The HYP is not in pain.

(vi) The HO is in pain.

Yet the above claims would seem to lead to a contradiction if we were also to subscribe to the following theses:

(1) *Subjectivity*: Necessarily, for any P-conscious state or experience *p*, there is a subject *S* for whom there is something it is like to be in that state or to undergo that experience.

(2) *Subjective unity:* Necessarily, at any particular time *t*, a P-conscious subject will have a single phenomenally unified field of consciousness; if the subject might be said to have multiple co-occurring (simultaneous) P-conscious states or experiences at *t*, then those states or experiences will necessarily be subsumed by a single unified phenomenal state or experience.

(3) *Consistency*: At any particular time *t*, it is impossible for any particular P-conscious state or experience *p* to both possess of, yet not possess of, a particular phenomenal property *F*.

(4) *Neurofunctional unity*: Necessarily, at any particular time *t*, any structurally-functionally intact human brain can support only one P-conscious subject.

(5) *Phenomenality of pain*: It is a necessary condition of pain that it has phenomenal properties.

I assume the first thesis to be self-evident, without any commitment to any particular metaphysical position on the nature of the subject of phenomenal states/experiences. As for

the other theses, while they seem to enjoy strong intuitive support, I will uphold the theoretical possibility of letting go of any of them should there be sufficient reasons to do so. Yet unless I indicate otherwise, in the following arguments, I will assume all of the above theses to be true.

So the problem is the following. It follows from the subjectivity thesis and the phenomenal thesis that the HO is either a P-conscious subject, or at least a P-conscious part of a subject. Yet according to the neurofunctional thesis, the HYP and the HO cannot constitute two distinct P-conscious subjects; at most, they can only be understood as constituting two distinct aspects or parts of a single P-conscious subject. Since, according to the consistency thesis, it is impossible for any single P-conscious state or experience to both possess of yet not possess of the phenomenal properties of pain, it would thus seem that the only way we could account for the HOP would be by assuming that subjects in such situations possess of two parallel streams of consciousness – yet this possibility is barred by the subjective unity thesis.

How shall we resolve this apparent dilemma? Since, in general, the case for hypnotic analgesia seems to be much better established than the case for the HOP, the skeptic is expected to (re-)attack claim (ii'). Yet since we have already ruled out the possibility that subjects in the HO situation are faking (etc.), it seems that the skeptic would have to deny either or both of the following two claims:

(vii) If the HO reports that she is in pain, then she believes that she is in pain.

(viii) If the HO believes that she is in pain, then she is in pain.

The intuition underlying (viii) is that which feeds incorrigibility theses about our subjective experiences. Indeed, it is very hard to imagine how one might possibly be wrong about his

own current pain. While some (e.g., Dennett, 1978) have argued that it is at least logically possible to have a false belief about one's own pain-states, for those (like me) who are skeptical about such a possibility, it would be convincing to see at least an empirical analogue of how such a false belief might be possible. A neurological example that comes to mind is Anton's syndrome (visual agnosia), a condition in which patients firmly deny having lost their vision despite obvious evidence to the contrary (their occipital lobes are damaged, they frequently suffer accidents due to bumping into walls and other objects, they fail on visual performance tests, etc.). Due to the fact that these patients persistently confabulate about their visual environment, this syndrome is not infrequently taken as a *par excellence* example of false beliefs about one's very subjective states.

Yet I fail to see how confabulation in visual agnosia might provide support for false beliefs about one's *internal* phenomenal states. Surely, persistent confabulation provides strong evidence regarding patients' false beliefs about the *external* visual stimulus state of affairs. Yet having a false perception in this sense is not equivalent to having no perception at all. Actually, it is very likely that patients claim to see exactly because they have phenomenal visual experiences – e.g., it has been suggested that such "confabulative behavior" might be an outcome of misinterpreting visual images due to a defective visual monitoring system, or perhaps due to false feedback arriving from another visual system (Maddula, 2009).

Consider another example: the pain experienced by patients with "phantom limbs". While pain perception in such patients might be thought of as non-veridical in the sense that they attribute pain to body parts which do not exist, no professional would nowadays doubt the psychologically genuine nature of the experiences of such patients (cf. Melzack, 1993).

Thus, arguments about confabulation and the non-veridicality of perception seem to miss the point: the issue is not whether things as represented match some particular state of affairs in

the external world, but whether one might believe that she was in pain when, as a (psychological) matter of fact, she wasn't. (Actually, veridicality arguments seem to fall off the mark in another important way as well: recall that, as opposed to patients suffering from Anton's syndrome, a central feature of the HO is that she can provide precise reports on the actual stimulus state of affairs.)

Perhaps a more promising route to undermining the reliability of HOs' reports, then, is to attack claim (vii) by questioning whether the HO really believes, after all, that she is experiencing pain. Besides arguments related to faking (which we have ruled out), there seem to be at least two further arguments the skeptic might come forward with. The first (and less sophisticated) proposal would be to assume that the HO has no beliefs at all because it (*sic*) is merely an "automaton". Surely, it might be argued, if covert reports amount to nothing over and above mere psychophysiological/behavioral reflexes, then the HO poses no serious threat to the notion of a single unified consciousness. Yet the automaton theory does not seem to go through, for subjects are well capable of providing covert reports on their pain in various modalities and in various alternative ways (e.g., by pressing buttons, providing verbal reports etc.). A passage from Binet (1890) illustrates well the point that "automatic reports" are not automatic in the sense of being reflexes:

> […] the manner in which the idea is expressed depends upon the attitude given to the anaesthetic hand. Thus, we ask the subject to think of the number 3. If he holds a pen in his hand he will write the figure 3. If he has no pen, and if before the experiment we have several times shaken the fingers of the insensible hand, the subject will raise his finger three times; the same will apply to the writs or to the movement of any other member. If the subject has a dynamometer in his hand he will press three distinct times upon this instrument. If the experimentalist himself assumes the initiative by raising the finger of the subject a certain number of times, the finger after having yielded three times to the impressed movement will stiffen, as if it thus wished to inform the experimentalist of the number that had been thought of. (p. 26)

Another (way more powerful) objection might come in the form of what we might call the argument from suffering. The starting point of this argument would be an important characteristic of the HO's report which I have not yet discussed: while the HO reports intensive pain, she concurrently reports not to be suffering or "bothered" by the pain. As a subject formulated this after a hypnotic session:

> The hidden part doesn't deal with pain. It looks at what is, and doesn't judge it. (Hilgard, 1977, p. 209)

The argument from suffering would thus go like this:

(ix) The HO is not suffering.

(x) Suffering is an essential feature of pain; i.e., if there is no suffering, there is no pain.

Therefore,

(vi*) *It is not the case that* the HO is in pain.

Of course, this is just what the skeptic would want to claim. Then, if she were to agree that it is impossible to have a false belief about one's own occurring pain, she could undermine claim (vii) by a reductio argument in the following way:

(viii) If the HO believes that she is in pain, then she is in pain.

(vi*) *It is not the case that* the HO is in pain.

Therefore,

(xi) The HO does not believe that she is in pain. [mod. toll., (viii), (vi*)]

(vii) If the HO reports that she is in pain, then she believes that she is in pain.

Therefore,

(iv*) *It is not the case that* the HO reports that she is in pain. [mod. toll., (vii), (xi)]

(iv) The HO reports that she is in pain.

Therefore,

(vii*) *It is not the case that* if the HO reports that she is in pain, then she believes that she is in pain. [neg. intr., (vii), (iv), (iv*)]

Yet the issue of credibility would now arise in a different form for the skeptic. For, on the one hand, the above argument entails the following claim (which is just what the skeptic wished to claim upon attacking (ii'):

(ii'*) *It is not the case that* if the HO reports that she is in pain, then the HO really is in pain. [neg. intr., (ii'), (vi), (vi*)]

On the other hand, in order to secure the claim that the HO is not suffering (claim (ix)), the skeptic would at least have to assume the following:

(xii) If the HO reports that she is not suffering, then the HO is not suffering.

Yet what could be the reason for discrediting the HO's report on pain (ii'*) if one assumed that the HO's report on suffering is credible (xii)? Since we have ruled out faking and false beliefs, the only possible option left would seem to be that the HO is using the term "pain" for a sensation/feeling which, though she associates with pain, she does not construe of as pain proper.

As it happens, there is strong empirical evidence that pain is a phenomenologically complex experience comprised of at least two components (Aydede, 2010; Grahek, 2007; Hilgard, 1977; Hilgard and Hilgard, 1994; Loeser and Melzack, 1999; Melzack, 1993; Melzack and

Casey, 1968; Rainville, Carrier, Hofbauer, Bushnell and Duncan, 1999). The so called sensory-discriminatory component relates to our capacity to identify and discriminate the location, duration, intensity and quality of pain, and is (primarily) supported by the somatosensory cortex in the brain. The so called affective-motivational component relates to the aversive (avoidance-related) nature of pain, and is (primarily) supported by the anterior cingulate cortex. Interestingly, these components are reported to dissociate in certain conditions, perhaps the most notable among these being pain asymbolia. In this neurological condition, patients fail to show any negative affective reaction even to severely noxious stimuli (they often smile, chat and laugh during the stimulation), yet they are well capable of detecting and discriminating the location, intensity and quality of pain.[10] Though Grahek (2007) argues that pain asymbolia is (might be) the only genuine empirical case in which the affective-motivational component is truly missing, he does not discuss hypnosis – as it seems, there *is* sound psychological/neuroscientific evidence that the suffering component of pain is significantly diminished in hypnotic analgesia (Rainville et al., 1999; Rainville, Duncan, Price, Carrier and Bushnell, 1997).

As expected, there is disagreement concerning the philosophical implication of such a dissociation between the two components of pain. While some argue that the mentioned cases have no bearing on the conceptual grounding of what it is for someone to be in pain, others believe that, in however rare cases, but if it is possible that a person should (a) identify an experience as a pain-experience, yet (b) not experience any "hurt", meanwhile (c) be capable of providing just as accurate reports on the stimulus state affairs as persons for whom those stimuli would be accompanied by suffering, then perhaps it is our concept of pain that needs

[10] This syndrome is not to be confused with congenital analgesia, a condition in which tactile sensing is intact, yet in which there is a complete shutdown of the pain system. Since persons inflicted with this disorder are totally incapable of feeling pain, they frequently suffer severe physical injuries, and it is not uncommon for them to die prematurely (cf. Grahek, 2007).

reevaluating (cf. Aydede, 2010; Dennett, 1978). Since my issue is with the skeptic, for the sake of argument, let us assume the more conservative view according to which suffering is an essential feature or component of (our concept) of pain. Of course, the skeptic need not deny that the HO experiences qualities associated with the sensory-discriminatory component of pain (let us call this $pain_{sensory}$). What she might posit, though, is that when the HO reports pain, she is actually referring to $pain_{sensory}$. This way, it would seem that the skeptic might actually cut the Gordian knot, for she might reconstruct the original argument without leading to a contradiction, and without assuming that the HO has false beliefs or is dishonest. Thus, the skeptic would start with the following:

(ii'') If the HO reports that she is in pain, then she is in $pain_{sensory}$.

(iv) The HO reports that she is in pain.

Therefore,

(vi') The HO is in $pain_{sensory}$.

The skeptic's job is not finished yet. For if subjects within the HO situation generally referred to $pain_{sensory}$ whenever they reported on "pain", then it would follow from (v) that the HYP is *not* in $pain_{sensory}$. Clearly, this would not help much in resolving our dilemma. Yet the skeptic might claim to have made progress under the assumption that hypnotized subjects understand very well the concept of pain. Hence, she might argue, when subjects (or, actually, the HYP) overtly report feeling no pain, what they (naturally) mean is that they are not suffering – i.e., that they are not experiencing $pain_{affective}$:

(i'') If the HYP reports that she is not in pain, then the HYP is not in $pain_{affective}$.

(iii) The HYP reports that she is not in pain.

Therefore,

(v') The HYP is not in $pain_{affective}$.

Yet while it is true that claims (v') and (vi') are compatible, this strategy will hardly work for the skeptic. For subjects within the HO paradigm commonly report on both $pain_{sensory}$ and $pain_{affective}$ – and, as it happens, the HYP explicitly reports *not* feeling any $pain_{sensory}$. So the headache for the skeptic comes in the form of the following statement:

(v'') The HYP is neither in $pain_{sensory}$, nor in $pain_{affective}$. .

In light of our theses, (v'') and (vi') are not compatible. Surely, we should not fall back on assuming that subjects are deceiving the experimenter when they claim not to be feeling any $pain_{sensory}$. So a last resort of the skeptic might be to assume that perhaps the HYP misunderstood or misinterpreted the questions concerning $pain_{sensory.}$ Yet this is highly improbable in light of two empirical studies. In one of these, low-hypnotizables were found to be (somewhat) capable of reducing the level of $pain_{affective}$ *without* a reduction in the reported level of $pain_{sensory}$ (Crawford, Knebel and Vandemia, 1998, *in* Kallio and Revonsuo, 2003). In the other study, high-hypnotizables could both reduce *and* enhance the level of $pain_{affect}$ *without* a change in $pain_{sensory}$ (Rainville et al., 1999).

The moral of the above is that the issue of whether the "suffering" component of pain is a necessary condition for a pain's qualifying as pain proper is orthogonal to our problem. For, under the assumed theses, the apparent contradiction between the HYP not feeling pain yet the HO feeling pain would seem to linger around under *any* understanding of pain as long as it was assumed that both the HYP and the HO are reporting on their P-conscious states or experiences of pain.

4. The dissociation of phenomenal consciousness and access-consciousness

4.1. Phenomenal consciousness without access-consciousness

In the last section, we examined whether, by providing alternative interpretations of "pain", we might dispose with the claim that if the HO reports that she is in pain, then the HO really is in pain (ii'). In this chapter, I will focus on the opposite strategy, and examine whether we might resolve the seemingly paradox nature of the HOP if we were to dispose with the following claim:

(i') If the HYP reports that she is not in pain, then the HYP really isn't in pain.

The main strategy I wish to focus on is based on "phenomenal overflow" arguments (Block, 2007, 2008), according to which, due to an informational bottleneck that severely constrains the amount of information that is cognitively accessible to us at any given particular time, it can happen that we have P-conscious experiences of which we cannot possibly be A-conscious (cf. also Bayne and Chalmers, 2003; Block, 1995/1997). The paradigmatic empirical study cited in support of this view is a classic experiment conducted by Sperling (1960) on "iconic memory". In this study, Sperling flashed an array of letters and numbers, e.g., in a typical configuration of 3 rows of 4 symbols each (see *Figure 2*.). In the first phase of the study, after a very brief (50 msec) exposure to the stimuli, the subjects had to recall as many symbols as they could ("whole report" condition). Using this method, Sperling established that the average number of symbols that can be retained from immediate memory[11] under such conditions is around 4 symbols (4.3 to be exact). In the second part of the study, subjects had to report on only one row of symbols following a tone that was

[11] Psychologists nowadays prefer the terms "short-term memory" or "working memory", but this is the term used by Sperling (1960) in his original paper.

presented half a second *after* the visual stimuli was flashed ("partial report" condition): if subjects heard a high tone, they had to report the symbols in the top row; if they heard a low tone, they had to report the symbols in the bottom row; and if they heard a medium-pitch tone, they had to report the symbols in the middle row. Now, since subjects did not know which tone they were going to hear after the array of symbols was flashed, if their average performance on *all* the rows (to be calculated by adding up the average performance on each individual row) was to be higher than in the whole report condition, then that would establish that there is more information available to the subjects than they can report on based on their immediate memory. Indeed, the average number of symbols available for report in the partial report condition was more than twice the number (9.1) observed in the immediate-recall trials.

Figure 2.

7	I	V	F
X	L	5	3
B	4	W	7

A typical array of symbols in Sperling's (1960) experiments on "iconic memory".

Is it possible that the HOP is an analogue of the Sperling situation, and that, thus, we might account for it in terms of limited access and phenomenal excess? In other words, could we reformulate claim (i') in the following way?

(i''') If the HYP reports that she is not in pain, then the HYP in not *A-conscious* of pain.

Block (1995/1997) argued that the HOP is at least a genuine candidate of P-consciousness without A-consciousness, at least insofar as the following is a possible interpretation of it:

> There is one system, the person, who has some sort of dissociation problem. There is P-conscious pain in there somewhere, but the person, himself or herself, does not have access to that pain, as shown by the failure to report it, and by the failure to use the information to escape the pain.[12] (p. 406)

I, for one, am deeply unconvinced that such a strategy can account for the HOP. First of all, it is an essential feature of the Sperling-type experiments that subjects assert seeing *more* than they can report on, or at least report having had *some* P-conscious experience of the stimuli (Block (2007) refers to this latter feature as "generic phenomenology"). Yet what could be the analogue of this in the case of the HOP? After all, it is not the case that subjects overtly report some "generic phenomenology" in relation to pain which they cannot specify or characterize or localize etc. – on the contrary, they firmly claim not to be feeling *any* pain at all.

Second, in the Sperling experiments, the very evidence of availability "overflowing" the capacity of "immediate memory" (accessibility) derives from the fact that the symbols in any single row can easily be reported *if* cued. This pattern of availability fits the description of access-disunity discussed in the *Section 1.3.*: while representational contents A and B might both be individually accessible, A&B are not jointly accessible (Bayne and Chalmers, 2003). Yet in the HOP, the HYP simply cannot access the pain. Of course, the problem is not that the HYP doesn't access the pain – since A-consciousness is a dispositional concept, "globally broadcast" content need not actually be accessed in order to qualify as accessible. Yet if pain were not even access*ible* for the HYP, than it is hard to see in what way the HYP might be said to be in pain. (Recall Levine's (2007, p. 514) already cited point in relation to this: "Phenomenal states/properties are not merely instantiated in the subject, but are experienced by the subject.")

[12] According to Block, a second possible interpretation of the phenomenon is that there are two different subjects within the same body. (He doesn't consider any further possibilities.)

Though Block (1995/1997) himself concedes that there is force to the claim that "if there is a P-conscious state in me that I don't have access to, then that state is not *mine* at all" (p. 406; italics in original), he nevertheless seems to believe that a "single-system" interpretation of the HOP might go through. Yet this seems hardly plausible in light of the fact that while the HYP claims not to be feeling any pain, at the same time, the HO is ardently reporting just the contrary. So it seems like we are dealing with two parallel cognitive systems – one of which is A-conscious of pain, the other of which isn't.

Then again, as mentioned in *Section 1.4.*, contrary to the split-brain case in which the left hemisphere and the right hemisphere are mutually ignorant of each other, the HOP seems to involve a relation of asymmetrical access between the HYP and the HO: while the HO is omniscient about things going on in the HYP, the latter is totally oblivious even to the very existence of the HO. This point about meta-access underscores that the HOP does not seem to be explainable away by appealing to informational bottlenecks; whatever it is by virtue of which the HYP is barred from gaining access to the pain seems unrelated to the capacity of the system.

Of course, an important question is whether P-consciousness might mirror/accompany the breakdowns in A-unity, or whether it might retain its unity in the face of A-dissociation. Though Bayne (2008) has meanwhile abandoned the idea, Bayne and Chalmers (2003) argued for the latter possibility, claiming that if representational content A in one of the cognitive systems is associated with phenomenal state P, meanwhile representational content B in another cognitive system is associated with phenomenal state Q, then even though there will be no cognitive system to which A&B will be jointly accessible, there will nevertheless

be something it is like for the subject to be in both P and Q at the same time.[13] Yet a clear problem with this view is that insofar as the HOP involves two cognitive systems, one of which is P-conscious of pain and one of which isn't, such a view in effect implies that there will be a P-conscious state which will subsumptively unify the apparently intramodally inconsistent P-states of the HYP and the HO.

As I mentioned in *Section 1.3.*, as opposed to Bayne and Chalmers, I am not so convinced that insofar as cases of A-dissociation involve phenomenal overflow, such cases involve dissociation only in a "shallow" sense. For the sake of argument, let us grant that it might be possible for a subject to be P-conscious without also being A-conscious. Perhaps this might just be the case in early infancy or in certain vegetative states insofar as persons in such cases might not at all be A-conscious. Yet a serious problem arises for phenomenal overflow models when we assume the occurrence of A-conscious states without allowing for those A-conscious states to be "zombie" states (of course, the proponent of the phenomenal overflow cannot allow for this unless she were to posit the obscure – if not outright incoherent – possibility of zombies who nevertheless have inaccessible P-conscious experiences). Now, arguably, the HYP is not a "zombie" – on the contrary, we might well assume that there must be *something* it is like for her to be in, say, a meta-(A)-conscious state in which she is fully A-conscious of the fact that she has no A-conscious experiences whatsoever of being in pain. And, arguably, this what-it-is-likeness is exactly what we would describe as a non-painful P-conscious state. Yet if this is so, then we are back to our original problem: two allegedly co-occurring P-conscious states (an accessible state of no-pain and an inaccessible state of pain) with intramodally inconsistent properties/content. So far, this is what I see as the implications of phenomenal overflow in a single system. Now add to this picture the obscure implication

[13] Note that this view does not entail representationalism as generally understood in the philosophy of mind. The claim is simply that *if* there is a mental representation that is P-conscious, then so and so.

of Bayne and Chalmers' subsumtive thesis that, in the case of A-dissociation, there will also be an in-principle-inaccessible P-conscious state unifying the painful and the non-painful state. What you get, then, are *two* in-principle-inaccessible P-conscious states from the point of view of the HYP: one of P-conscious pain, and another of a unified P-conscious hybrid of pain/no-pain. Shallow or non-shallow – adhere to this view, and you will have two dissociations with intramodal inconsistency on top.

Let me point out what I see as the main fallacy in both Block's and Bayne and Chalmers' interpretation of the implications of Sperling-type experiments. What studies on "iconic memory" establish is that *even if* all the information contained in a P-conscious state at time t_1 could be said to be also A-conscious (i.e., potentially available for further processing) at t_1, due to constraints on cognitive processing, only a certain amount of the available information at t_1 can actually be accessed (i.e., further processed or "consumed") at time t_2. Yet this in no way implies that, at any given time, however much we might be ready for it (e.g., by focusing our attention to that feature/aspect), there might nevertheless be certain particular features/aspects of our P-conscious states/experiences which we cannot possibly (in principle) have access to.

So if the "breakdown" of access unity simply refers to the seeming fact that we cannot possibly access all the information contained in a phenomenal state at any given moment, then I agree: this only implies a dissociation of consciousness "in a relatively shallow sense." On the other hand, if one were to posit the possibility of persisting phenomenal pain without that pain's being even potentially available for access, then I believe one would in effect be implying a dissociation in P-consciousness in quite a "deep" sense.

4.2. Access-consciousness without phenomenal consciousness

In the previous section, I argued that reinterpreting the experience of the HYP in the spirit of phenomenal overflow arguments would not cut much ice if one's intention were to refrain from letting go of our consistency and/or unity theses. In this section, I wish to reorient the focus on the HO once again, and pursue whether assuming the opposite of phenomenal overflow, viz. the possibility of A-consciousness without P-consciousness might help us in resolving the apparent contradiction between the HYP's and the HO's reports. More specifically, I am interested in whether we might make progress by assuming the following:

> (ii''') If the HO reports that she is in pain, then the HO is A-conscious (but not P-conscious) of pain.

In his original paper on the conceptual distinction between A-consciousness and P-consciousness, Block (1995/1997) emphasized that, at least under his construal of the concept, blindsight patients are not A-conscious of things within their blind field. A convincing support for this claim is that even if a blindsight patient were thirsty, she would still not reach for a glass of water in her blind field. In other words, even if the glass of water were detected by her collicular system, thus activating a mental representation which might positively bias her forced-choice *guesses* concerning a very limited set of some very basic features of the glass of water, the patient would nevertheless be incapable of using this representation for the purposes of reasoning, verbal report or the rational control of behavior.

Yet Block conceptualized of what he dubbed "super-blindsight", a phenomenon in which a patient would be A-conscious of the representations activated by visual projections to her blind field without having any visual experiences proper of the projected stimuli. In Block's words (1995/1997, p. 385):

> Visual information from his blind field simply pops into his thoughts in the way that solutions to problems we've been worrying about pop into our thoughts, or in the way some people just know the time or which way is north without having any perceptual experience of it. The super-blindsighter himself contrasts what it is like to know visually about an X in his blind field and an X in his sighted field. There is something it is like to experience the latter, but not the former, he says."

Though Block left the question open as to whether there might be any actual cases of A-consciousness without P-consciousness, he exclaimed that "the (apparent) nonexistence of super-blindsight is a striking fact" (p. 386), ultimately concluding that such cases "do not appear to exist" (p. 402). Block might well be right concerning blindsight patients, but I am more interested in whether there might be any cases of "super-blindsight" if the term were construed of in a more general sense, viz., as applying to any case in which one might gain access to a representation of an object or a perceptual state without at the same time having any P-experience of the sensory modality through which the object is presented, or with which the perceptual state is generally associated. Construed of in this way, the following phenomenon, observed in anesthesia (analgesia), would seem to qualify as a paradigmatic case of "super-blindsight":

> Things placed in the hand were not felt, but *thought* of […]. A key, a knife, placed in the hand occasioned *ideas* of a key or a knife, but the hand felt nothing. Similarly the subject *thought* of the number 3, 6, etc., if the hand or finger was bent three or six times by the operator, or if he stroked it three, six, etc., times." (James, 1890/1983, p. 202; italics in original)

Is it possible that the HO might be a "very limited partial zombie" (Block, 1995/1997, p. 385) of this kind, with thoughts/ideas of pain "popping" into her mind without any feeling of pain? Consider the following retrospective report of a subject:

> Part of me knows the pain is there, but I'm not sure I *feel* it. The hypnotized part *doesn't* feel it, but I'm not sure if the hypnotized part may have known it was there but didn't say it. (Hilgard et al., 1975, p. 286; italics in original)

This would seem to fit the picture: there is no feeling of pain, just "pure" knowledge about it. Then again, insofar as the HYP was aware of this knowledge all along, this case is not a

paradigmatic example of the HOP. Sure enough, it is an exciting example of just the kind of "super-blindsight" we have been discussing – but if the subject knew about her knowing, then there is no contradiction involved. But the truly baffling characteristic of many HO cases is that the HYP is totally ignorant about both the HO in general (including its reporting activities), and the pain in particular. In the words of an experimental participant:

> It's strange, because I didn't knowingly feel the pain reported, and (in automatic writing) I didn't knowingly know I was writing. (Hilgard, 1977, p. 210)

So the problem is the following: insofar as the HYP is totally oblivious of any "stand-ins" (like thoughts or cross-modal sensory experiences) of pain, our dilemma concerning the apparently paradoxical nature of HOP would seem to linger on. For even if the HO were to acquire knowledge about what it considers as pain without any sensory experiences of pain, insofar as we were to assume that the HO is P-conscious of such "knowledge", our problem would have simply shifted from how one might both *feel* pain yet not feel pain to the problem of how one might possibly both *think* that she is in pain yet not think that she is in pain. While in *Section 3*. we only concluded that the dilemmatic nature of the HOP will remain under any interpretation of pain insofar as we assume the HO to be P-conscious of pain, it is now clear that our real issue has little to do with the feeling of pain per se: the dilemmatic nature of the HOP will remain under any interpretation of what the HO reports on insofar as there is a single mental state or experience of which the HO is P-conscious, yet of which the HYP is not.

Hence, while certain cases of hypnotic analgesia might well constitute good empirical examples of super-blindsight, positing HOs to be "very limited partial zombies" will not resolve our real problem; if we are going to posit zombiehood, we will have to go all the way down that road. Interestingly, a gut-rejection of the zombie interpretation of the HO seems to be the converging point of practically all parties interested in the HOP. Yet while

philosophers like Block (1995/1997) or Bayne (2007) argue that the HO must be (or is at least most probably) P-conscious, psychologists like Bowers (e.g., 1994) or Dienes and Perner (2007) seem to imply that the HO might not even qualify as (fully) A-conscious.

Block (1995/1997) proposed two reasons why we might believe that the HO is P-conscious. The first is based on the common observation that, in hypnotic analgesia, certain psychophysiological markers of pain (e.g., heart rate and blood pressure) manifest the same level and pattern of activation as under P-conscious pain. Yet such a reasoning seems to me neither empirically, nor philosophically very promising. First, psychophysiological markers of pain do not covary with the intensity of stimulation in the same lawful way as shown by self-reports of pain (Hilgard, 1969, *in* Kihlstrom et al., 2000); thus, heart rate and blood pressure would hardly seem like sufficient indicators of P-conscious pain. Second, it would seem philosophically futile to argue against zombiehood based on somatic markers – after all, it would only seem to follow from zombie arguments that if P-consciousness is contingent, then it might very well dissociate from the physical states/processes it is generally associated with.

According to Block (and Bayne; and I suspect most non-psychologists in general), a second reason to believe that the HO is P-conscious is that she often describes the pain as "excruciating". We might add to this the further important point that subjects recount even retrospectively of their HO experiences as having been P-conscious. While Hilgard (1977) – the modern re-discoverer of the HOP – was well aware of this, he nevertheless believed that, up till the point of the retrospective accounts, "the observing part was hitherto not in awareness" (p. 185), claiming that the HO is merely "a metaphor for something happening at the 'intellectual' level but not available to the consciousness of the hypnotized person" (p. 188). Though arguments for this position are hard to find in Hilgard's work (something

Bayne (2007) has also commented on), I believe it is not at all as absurd or skeptical a position as it might at first seem.

First of all, one might argue that if letting go of at least one of our theses were unavoidable in light of the HOP, it would still be less counterintuitive to assume the empirical viability of our possibly reporting unconsciously on pain (or the bodily/cognitive states/processes associated with pain, if you like), than to assume the empirical possibility of a genuine dissociation of P-consciousness (including cases of phenomenal overflow). Second, recall the case of Swamp Mary, an imaginary neuroscientist of vision imprisoned in a black and white room just like Jackson's (1982) famous Mary (who, before leaving this room, has no knowledge of *what it is like* to see colors), with the only difference that, before her release, Swamp Mary is struck by a bolt of lightning that miraculously rearranges her brain in such a way that her brain state becomes identical to the state she would have gone into after seeing, say, a rose for the first time (Dennett, 2007, *in* Alter, 2008). Arguably, there is nothing conceptually inconceivable about such a scenario; and what the HOP suggests (if Hilgard's view is correct) is that such cases might very well be empirically possible – without assuming any "miracles". As an experimental participant formulated this in relation to the HOP:

> Maybe the tones register in your memory and skip going through the conscious part of your brain.[14] (Hilgard, 1977, p. 210)

There is an interesting and important implication of the "Swamp Mary interpretation" of the HOP which – to the best of my knowledge – has not yet been discussed in the literature. What I have in mind is the implicit relation between the HOP and "lucid dreaming". Lucid dreamers are people who claim to be perfectly aware during their dreams that they are dreaming, without this awareness interrupting their sleep-state or their very dreaming. Since

[14] This subject also participated in a session in which hypnotic deafness was suggested, thus, the "registered" tones refer to sounds allegedly not heard by the HYP.

subjective reports of lucid dreaming were originally met with great skepticism, researchers took to the task of finding evidence (beyond subjective reports) for the existence of such dream-states. In a famous series of studies, researchers asked well-trained lucid dreamers to perform some prearranged signal (e.g., an eye-movement or a fist-clench) during their lucid dreams. Such signals were then matched both with polysomnographic measures and subjects' retrospective accounts of whether – and if yes, when and how many times – they had lucid dreams during their sleep that night. As it turned out, there was a correspondence between the reported and the observed signals, all cases of which occurred during unambiguous REM sleep (the sleep-phase associated with dreaming) (cf. LaBerge, 1990).

Here's the issue. Evidence on lucid dreaming is now widely considered to be well-established, and few professionals would posit that subjects in the mentioned studies were only unconsciously signaling to the experimenters, becoming P-conscious of their dreams only retrospectively (viz., upon awakening). So the question is the following: if, based on a fit between the behavioral, psychophysiological and subjective data, one were willing to grant that lucid dreamers really have P-conscious lucid dreams, then why deny, in the face of just such a fit between just such kinds of empirical data in the case of the HOP, that the HO is P-conscious? Or to approach the issue from the other direction: if we might have reason to believe that, despite claims of memories to the contrary, the HO is not P-conscious, then shouldn't we perhaps re-evaluate our position on lucid dreaming (or dreaming in general, for that sake) as well?

A common argument in favor of ascribing P-consciousness to lucid dreamers is based on the observation that subjects clearly manifest goal-directed behavior. In the spirit of such reasoning, one might argue about the HO in the following way:

(xiii) It is a necessary condition of goal-directed behavior that a subject be P-conscious.

(xiv) The HO manifests goal-directed behavior

Therefore,

(xv) The HO must be P-conscious.

While the argument admittedly has some intuitive appeal, the skeptic will clearly want more than mere "folk" intuition in relation to the first assumption. More specifically, he might ask why the following claim should not suffice for our purposes:

(xiii') It is a necessary condition of goal-directed behavior that a subject be A-conscious.

Based on Shallice's (1997) argument in favor of the view that the right-hemisphere of split-brain patients is (most probably) P-conscious, one might argue for (xiii) by claiming the following:

(xvi) If goal-directed behavior were possible in the absence of P-consciousness, then P-consciousness would be epiphenomenal.

(xvii) P-consciousness is not epiphenomenal.

The above claims, if true, would clearly entail (xiii). The problem is that both claims seem to beg the question on epiphenomenalism, especially if the term is to be understood in the psycho-functional sense here.[15] For, as I will soon discuss the issue, even if one adhered to (xvii), it would in no way follow that goal-directed behavior was contingent on

[15] Psycho-functional epiphenomenalism is a position according to which P-consciousness has no role to play in our psychological functioning. This position is neutral concerning the issue of metaphyisical epiphenomenalism insofar as it does not posit (and neither does it deny) that P-consciousness is causally inert with respect to the physical.

P-consciousness (or A-consciousness, for that sake). Of course, it should be noted that epiphenomenalism *per se* is not really related to our issue concerning whether the HO is P-conscious. For, e.g., even if P-consciousness had no causal role to play in the bringing about of goal-directed actions, one might easily admit to (xiii) if she were to admit both to (xiii') and to the following empirical claim:

(xviii) Any A-conscious state/experience is also P-conscious.

Of course, (xviii) is exactly what the zombist would deny, but let us put this issue aside for now. The second reason why the issue of epiphenomenalism is orthogonal to our topic is that, independently of whether one adhered to (xvii) or (xviii), one might argue (contra (xiii')) that not even A-consciousness is a necessary condition of goal-directed behavior. For example, Bowers (1994; Bowers and Davidson, 1991; Miller and Bowers, 1993) argues that certain non-conscious cognitive subsystems of behavioral control might, under certain special circumstances (e.g., hypnosis), become automatically activated and dissociate from the central executive system the task of which (under normal conditions) is to initiate/inhibit, coordinate, and oversee the functioning of the various subsystems. According to this theory of "dissociated control", then, there is nothing contradictory in the notion of an unintentional goal-directed action; behavior might well serve important purposes without its being initiated or maintained voluntarily. Of course, insofar as voluntary (intentional) action is understood as involving the central control and monitoring of various subsystems, the dissociated control theory entails (at the minimum) the following claim:

(xiii'') It is a necessary condition of voluntary (intentional) behavior that a subject be A-conscious.

Now, if goal-directed behavior can be automatic (non-A-conscious), then, clearly, even if one adhered to (xviii), in order to secure (xv) based on (xiii''), one would also have to assume the following:

(xiv') The HO manifests voluntary (intentional) behavior.

Of course, this is just what the dissociated control theory denies, according to which there is no executive initiative, directed allocation of attention, or controlled effort observable in hypnotic phenomena. Admittedly, there is some evidence in favor of this view (for a good general overview on automatic processes in hypnosis, cf. Lynn and Rhue, 1991). Yet there is also strong evidence to the contrary, viz. that subjects under hypnosis exhibit high-level executive functioning. Actually, there is now evidence that highly hypnotizable subjects can even voluntarily inhibit automatic processes which were until recently thought of as "mandatory" (i.e., unstoppable once activated). For example, in the classic Stroop test, in which subjects are asked to name the color of the ink in which a word appears, due to automatic lexical processing, subjects usually take more time to answer and commit more errors when the word presented itself denotes a color name (e.g., the word "blue" written in red color). Yet while this interference effect is very robust and non-eliminable under normal circumstances, when high-hypnotizables are given suggestions to hinder the lexical processing, amazingly, the Stroop interference is eliminated (Raz, Shapiro, Fan and Posner, 2002; cf. also Dienes and Perner, 2007).

Yet even if subjects under hypnosis can, in general, engage in high-level executive functioning, the question is whether the HO engages in any such cognition. Bayne (2007) argues that yes, she does:

> […] hidden observer behaviours appear to involve more cognitive sophistication and flexibility than those seen in those syndromes in which we are most inclined to invoke zombie systems. (p. 98)

Note, though, that Bayne conceives of "zombie systems" as involving those cases in which the goal-directed behavior of subjects is inflexible and unsophisticated – i.e., in cases in which there is no clear sign of executive functioning. Yet insofar as such cases need not involve A-consciousness (information might feed into the subsystems without being globally broadcast), I am not convinced we should construe of such automatic processes as involving "zombie systems" in the first place. After all, we don't refer to the automatic lexical processes activated during the Stroop task as involving "zombiehood", either. So, in my view, the question of zombiehood, at least in the philosophically/psychologically relevant sense, arises only at the A-conscious level.

Now, if one subscribed (as Bayne clearly does) to the view that the HO manifests voluntary behavior (xiv') *and* to the view that A-consciousness entails P-consciousness (xviii), then the following would clearly follow:

(xiii''') It is a necessary condition of voluntary (intentional) behavior that a subject be P-conscious.

Bayne hismself doesn't provide any independent argument for (xviii) – the very claim the skeptic/zombist will tend to challenge (of course, admittedly, (xviii) is notoriously hard to argue for on independent grounds). Instead, Bayne tackles the issue indirectly, arguing against (xiii'') in the following way:

> The answer given by the zombie model is prima facie implausible: if – as I have suggested – hidden observer behaviours are reports of bodily and perceptual states, then we ought to treat them as we treat other reports of perceptual and bodily states – viz., as representations of the subject's conscious states. (p. 97)

It is clear from the context of the above passage that what Bayne has in mind is P-consciousness. For the sake of argument, let us grant that Bayne is correct insofar as reports of pain (or at least $pain_{sensory}$) are reports of perceptual and bodily states. Yet how does

it follow from this that we should treat such reports as representations of the subject's (P-)conscious states just because that's how we *usually* go about interpreting such reports? After all, the HOP is a rather peculiar and uncommon phenomenon; so perhaps our common assumptions/intuitions concerning the relation between sophisticated reporting and P-consciousness are simply not applicable to cases like the HOP.

Actually, insofar as reports of pains might be understood as reports of bodily states, contra Bayne, one might perhaps argue in the following way (adapted from Kihlstrom et al.'s (2000) argument for unconscious emotions):

(xix) $\text{Pain}_{(\text{sensory})}$ is the perception of bodily states.

(xx) Perception can be unconscious.

Therefore,

(xxi) $\text{Pain}_{(\text{sensory})}$ can be unconscious.

As support for (xx), one might make reference to the mounting evidence on "implicit perception", which we might define as "the effect on the subject's experience, thought, or action of an object in the current stimulus environment in the absence of, or independent of, conscious perception of that event" (Kihlstrom, in press). Conscious or explicit perception in this context would relate to perception in which the perceptual contents were available for reasoning, verbal report, and the rational control of behavior. But then the problem is the following: If "unconscious" is merely understood as a synonym for "implicit", then the above argument would seem not to bear on our original issue, viz., whether the HO might be A-conscious of perceptual content without also being P-conscious. On the other side, if "unconscious" is understood as applying to all cases in which a subject is P-unconscious, then evidence on implicit perception would seem like no support for claim (xx) insofar as implicit perception does not involve A-consciousness.

Then again, psychologists and neuroscientists are now claiming that a growing body of evidence points to the direction that many cognitive processes formerly thought of as paradigmatically involving P-consciousness might actually occur without it. Perhaps the most striking examples of such processes are those of unconscious working memory (Hassin, 2005) and unconscious meta-cognition (Glaser and Kihlstrom, 2005). The former (if possible) is striking because working memory constitutes the very "workspace" where the flexible planning and control of behavior takes place; the latter (if possible) is striking because meta-cognition involves (among others) self-monitoring. Of course, one might ask how we can be sure that subjects really weren't P-conscious all along (perhaps subjects were simply not A-conscious of their being engaged in meta-cognition?). Yet even granted that such high-level unconscious cognition is possible, there remains a crucial difference between such cases and the HOP. For while, in the mentioned cases, subjects never have any (meta-)access to their being engaged in any high-level cognitive processing, subjects in the HO case gain full access at least to the memory traces of both prior pain-experiences (or the bodily/cognitive states retrospectively interpreted/remembered as phenomenal pain) and the reporting activities of the HO.

So, however sophisticated certain P-unconscious processes might be, admittedly, the HO is capable of pulling off a feat which subjects engaged in various other purportedly high-level unconscious cognition do not seem to be capable of. This might seem to tilt the balance toward the view that HOs might, after all, be P-conscious. Then again, one might still assume that the retrospective reportability of the HO's experiences is mediated by the simple fact that representations of pain were more deeply processed and better reinforced in memory as a function of the HO's gaining access to them for the purposes of "automatic reporting" during the hypnotic session.

I don't think we can call any shots here. Surely, as counter-intuitive as the notion of an unconscious $pain_{(sensory)}$ might be, if one is willing to grant that even working memory, meta-cognition and the voluntary control of flexible behavior might occur (P-)*un*consciously, then perhaps the idea of unconscious reporting on unconscious pain is not that obscure as it might at first seem. On the other hand, admittedly, it is hard to see why a HO that can provide rather sophisticated "online" reports on her bodily/cognitive/perceptual states should be denied P-consciousness if one had no second thoughts on ascribing P-consciousness to the HYP under similar circumstances.

5. The breakdown of temporal continuity

For those who take the reports of both the HYP and the HO to be fully credible, yet who do not believe in the empirical possibility of A-consciousness and P-consciousness coming apart, yet who are still not willing to let go of any of our theses, there seems to be one option left for making sense of the apparently contradictory reports of the HYP and the HO. This strategy involves challenging the truly simultaneous nature of the HYP's and the HO's reports, thus introducing time-indexicals to our original claims in the following way:

(iii') The HYP reports that she is not in pain at time t_1, t_3, …, t_i.

(iv') The HO reports that she is in pain at time t_2, t_4, …, t_j.

Bayne (2007) calls this the "switching model" (SM), for the basic idea is that there is only one (part or the) subject present at any time during the HO experiment who, upon the suggestion of the hypnotist, shifts or switches her attention to and away from the noxious stimulation. Of course, the hypnotist never explicitly suggested to the subject that she should shift her attention; on the contrary, in the original HO experiments, the hypnotist only referred to ongoing information processing of which the subject might be unaware (cf. Hilgard, 1977, p. 186). Yet proponents of the SM, building on the notion of demand characteristics, posit that subjects might have misunderstood or misinterpreted such suggestions as implicit requests to focus on the noxious stimulation. Of course, one need not fall back on assuming subjects to be faking (etc.). On the contrary, modern theories of role-enactment allow that subjects within the HO situation might themselves come to construe of their oscillating experiences of pain as simultaneous experiences arising from a genuine dissociation of consciousness (Spanos and Hewitt, 1980). Before I elaborate on this latter point, though, let me first make a couple of comments on the demand-compliance issue.

Recall that subjects are capable of reporting on the location, duration, intensity and quality of pain even if the HO is suggested only *after* the physical stimulation is terminated. So appealing to demand characteristics and/or misinterpreted suggestions will not do if subjects really do "register" information spontaneously. Then again, I would argue that *even if* one believed that HOs might provide retrospective veridical reports without any information on pain actually having been "registered" (due to evident contextual cues etc.), this would still leave open the possibility that, akin to lucid dreaming, perhaps subjects might have nevertheless had two parallel fantasy-experiences simultaneously. Yet let us not pursue this issue further, and grant that subjects (or their cognitive system) really do engage in spontaneous registering of the painful stimuli.

Interestingly, Bayne (2007), while himself a proponent of the SM, believes that evidence on the spontaneous registration of pain militates against the SM. Yet I'm not sure this is necessarily so. After all, the proponent of the SM might well rest his case concerning situational cues and demands, misinterpreted suggestions, etc., and nevertheless propose that the spontaneous registry of pain during hypnotic analgesia is effectuated by subjects' spontaneously shifting their attention to and away from the painful stimuli. So, as I see it, the question of switching has not much to do with demand characteristics or veridicality. Rather, the issue is whether we might account for the HOP by assuming that subjects have oscillating experiences.

Let us now return to the issue that subjects might themselves construe of their alternating experiences as simultaneous. Though a bit baffling in itself, there is nothing very controversial in the assumption that subjects might retrospectively (post-hypnotically) construe of prior alternating experiences as having occurred simultaneously. Of course, the problem is that subjects don't appear to experience any switching of their conscious

experiences during the hypnotic session, either – generally, both the HYP and the HO claim that their experiences are continuous. As it happens, no proponent of the SM (at least nobody I know of) has so far come forward with an account of how it might be possible for persons to have the "hyper-illusion" that their alternating, discontinuous experiences constitute parallel continuous experiences. Ironically, perhaps it was Hilgard (1977) – a fierce critic of the SM – who came closest to a possible explanation. As he noted in a passage on the simultaneous vs. the switching interpretation of the HOP,

> the classification of the two behaviors as dissociated applies to both interpretations. This is true because the *amnesic component* common to both convinces the subject that the two reports represent a pair of ongoing experiences." (p. 237; italics added)

While Hilgard didn't elaborate on the idea, I think we can get a hunch of what he probably had in mind if we consider the following case, also recounted by Hilgard:

> This subject remaining hypnotized had just had the tourniquet removed following the release of hypnotic analgesia in a repetition of an ischemic pain session. When asked whether the arm was comfortable, he replied that it was comfortable enough, but that there was some throbbing in the arm, gradually decreasing. With the hypnotist's hand on his shoulder for the hidden report, he described the concealed pain he had experienced in analgesia. With the hand removed and the subject amnesic for the covert experience, he reported that there had been a 'step decrease' in the throbbing of his arm. While the hidden observer inquiry was going on, the overtly responding part was unaware of the throbbing. The hypnotist placed his hand on the shoulder again, and the subject reported that the throbbing had decreased continuously and that the previous report of a sudden change was a mistake. The hypnotist lifted his arm from the shoulder, and the subject repeated: 'That's funny, there's been another step decrease in the throbbing.'" (pp. 238-9)

Of course, the above case is not a paradigmatic example of the HOP insofar as the HYP was not analgesic; in this case, all that was needed for the HO to experience the gradual decrease of the throbbing was that the amnesia of the HYP be lifted. *Mutatis mutandis*, here's my switching account of the paradigmatic HOP then. At time t_1, the subject (HYP) feels no pain. At t_2, she shifts her attention to the painful stimuli and feels "excruciating" pain. Yet at t_3,

upon switching back to the HYP, not only does she shift her attention away from the painful stimuli, but she is also struck by total amnesia regarding all that happened at (during) t_2. Now, since the subject is sure (quite righty) at t_3 that there were no temporal "gaps" in her stream of consciousness since t_1, yet since the HYP hasn't the faintest memory of anything having happened since t_1, the HYP mistakenly construes of t_3 as t_2, resulting in a *mis*conception of a continuous experience of analgesia. Then, at t_4, upon switching once again to "HO mode", the subject is oblivious of the fact that she wasn't in pain just an instance ago (at t_3). A possible explanation for this amnesia is that she accepts and builds on the time-construal of the HYP – after all, she *is* the same subject. So since the subject believed at t_3 (in "HYP mode") that it is only t_2, it seems logical that she should believe at t_4 (in "HO mode") that it is only t_3. Since, in "HO mode", the subject remembers that she was also in pain at t_2, the HO will also come be construed of as having a continuous stream of consciousness (in this case, flowing from t_2 through t_3).

We might say, then, that at any time t_i, the HYP will think that the time is $t_{(i+1)/2}$, while at any time t_j, the HO will think that the time is $t_{(j/2)+1}$. Now, recall that the HO has access to the HYP's states. Of course, since according to the SM, at no time can the HO and the HYP be present simultaneously, what the HO will have access to at any time t_j are the memory traces of the HYP leading up to t_{j-1}. Let us assume, then, that the HO is quizzed at t_{100}. According to our formula, the subjective time for HO at this time instance will be t_{51}, and based upon her memory traces, she will report having had a persisting P-conscious pain during the time span $t_{2\text{-}50}$. Of course, at t_{100} the HO also has full access to the memory traces of the HYP up till t_{99}, which, according to our other formula, was perceived by the HYP as t_{50}. Since the HO is well aware of the HYP's perception of an uninterrupted stream of consciousness up till that point as well, she will in fact report that, during the time span of t_2-t_{50}, the HYP and the HO were both continuously P-conscious!

As bizarre as the proposed model might seem, it does have the virtue of providing at least a tentative explanation of why/how it might seem to subjects that they have (had) two simultaneous experiences, without necessitating the abandonment of any of our cherished theses. Unfortunately, the model also has its serious drawbacks, and, ultimately, I don't believe there is any empirical evidence supporting it.

A first problem is clearly that, according to the model, one might have constant radical switches in her conscious experiences without any awareness, knowledge or recollection whatsoever of such experiential oscillation occurring or having occurred. Of course, there are plenty of examples of a person's entering into an altered state of consciousness without at the same time being aware (meta-conscious) of any alterations in experience. Such examples might include dozing off into light sleep, entering a state of trance, becoming intoxicated by a psychoactive substance, etc. Yet while in these cases we are well aware of the changes retrospectively, according to the SM, we are forever doomed to oblivion in the HO scenario.

A second (related) problem is that, as we saw, the model implies that one might easily be the victim of a hyper-illusion concerning the temporal continuity of one's own stream of consciousness. Yet perhaps it is not only the case that continuity breaks down in the HOP, but that continuity is an illusion altogether? Admittedly, this would be quite an unwanted implication of the model – after all, it seems a most fundamental feature of conscious states that they evolve through time. Yet the possibility of illusory continuity raises a perhaps even more disturbing possibility, viz. that our conception of persisting subjects of experience might be illusory as well. Naturally, this is no issue to be taken lightheartedly – e.g., if there are no persisting subjects, then the counterfactual hypotheticals we base our ethical judgments on are apparently nonsensical, in effect threatening the very notion of moral responsibility.

The good news is that, as noted, the SM does not seem to enjoy any empirical support. True, the single study that addressed the issue of switching directly found that, in 15 out of 16 cases, "overt" and "covert" reports of the subject were separated from each other by at least 500 milliseconds (Spanos and Hewitt, 1980). Yet one might have reservations concerning the validity and/or generalizability of these results. First, the study involved the measurement of behavioral output in differing modalities, so successive reporting in this sense would constitute no decisive evidence against simultaneous experiences. Of course, one might argue for an opposite point as well, viz., that even if there were behavioral evidence that overt and covert reporting do occur simultaneously, this would still not undermine the SM insofar as the conscious states underlying simultaneous behavior need not themselves be simultaneous (cf. Bayne, 2007). Yet even if one conceded that neither subjective experience/reports, nor behavioral measures provide decisive evidence against switching, one would at least expect (neuro)psychophysiologial evidence for the purported attentional switching (at least insofar as one assumes, as I do, that there is, at a minimum, psychophysical correlation between our conscious experiences and the states/processes of the brain) – evidence which is totally lacking.

Take for example the case of binocular rivalry, a phenomenon that occurs when a subject views two incongruent stimuli (e.g., a grated red image and a grated blue image) projected independently to each eye. In such cases, instead of having a single percept of only one of the stimuli, or of a fused image of the two stimuli, the subject experiences a constant switching back and forth between two percepts. Of course, in the case of binocular rivalry, subjects are well aware of the switches taking place – but the point is that there are also robust neural indicators of the switches (Tononi, Srinivasan, Russell and Edelman, 1998). Or perhaps the issue of meta-cognition in the case of binocular rivalry is not irrelevant at all, and the robust neural indicators might merely be reflections of subjects' noticing/signaling the switches?

This is a complicated matter, but the general point remains: if subjects' P-conscious experiences can really flicker between two incongruent states, then we would expect to find *some* differences between those states. (Note that simply assuming that such flickering has not been detected so far due to its occurring only sporadically is a non-starter: as mentioned, subjects can provide precise retrospective descriptions of patterned stimuli even when hidden observer suggestions are given after the termination of the stimulation and most subjects do not have "step-wise" experiences. So switches must be quite frequent and rapid.)

There is another important feature of binocular rivalry which seems to militate against the SM. For while, true, the rate of spontaneous switching in binocular rivalry can be biased voluntarily, there is no evidence that one might inhibit the occurrence of such switching – or the awareness of such switching if you like – altogether (Cosmelli and Thompson, 2007). On the contrary, there is firm evidence that, in the case of highly hypnotizable persons, such switches actually occur with a higher frequency, faster latency, and greater intensity than in the case of low-hypnotizables (De Pascalis, 2007).

So I, for one, am deeply skeptical of the empirical plausibility of the SM in light of the fact that there seem to be no evidence supporting it, and the closest examples of perceptual/attentional flickering we can think of differ significantly from the proposed account at both the experiential and at the objectively detectable level. The empirical issues notwithstanding, as discussed, the model also has its conceptual/theoretical drawbacks. Yet even if one were to concede to the potential implications of the model, it is at least questionable whether the model is preferable to a model of dissociated consciousness. After all, insofar as the model admits to psychologically genuine experiences of dissociation, it is not clear what the philosophical import of holding on to the notion of unity is in the face of potentially having to abandon the notion of a temporally continuous consciousness.

6. Summary and conclusions

Arguably, the HOP is one of the most intriguing and most controversial phenomena that hypnosis research has uncovered. The phenomenon is intriguing, for it provides further evidence of the amazing complexity and power of the human mind. Then again, the phenomenon is highly controversial, for it seems to be in conflict with our "folk" intuitions about the unity of conscious experience, as well as with all current models of consciousness within the fields of psychology and the cognitive sciences.

The primary question underlying this thesis was whether the HOP might constitute a genuine case of dissociated consciousness. Yet as desirable as it may be to know the answer to this notoriously challenging question, the primary objective of this thesis was not to arrive at a firm conclusion. Rather, the main goal was to show that, however we might attempt to avoid theoretical and/or conceptual conflict, once the details of the phenomenon are exposed, we shall hardly find an account of the phenomenon devoid of unwanted implications.

My method of inquiry concerning whether the HOP might constitute a case of a dissociation of consciousness was indirect. Adhering to the principle of the "presumption of innocence", my methodological stance was that we should not assume a breakdown in the unity of consciousness until we have not considered all other possible explanations. Thus, I started out by considering whether we might discredit or otherwise explain away the empirical data available on hypnotic analgesia (*Section 2.1*.) and/or the "hidden observer" (*Section 2.2*.), but following a rather detailed exposition of the most important issues related to these questions, I concluded that the evidence and the arguments in favor of the HOP are overwhelming. As a second attempt at explaining away the phenomenon, I examined whether we might dissolve the apparently contradictory nature of the HOP by attacking and/or reinterpreting claims concerning the veridicality of pain-reports (*Section 3*.); here, I arrived at the conclusion that,

as central as they might at first have seemed, the various philosophical questions and controversies surrounding the issue of pain are orthogonal to the dilemma posed by the HOP.

Seeing that there was no escape, in *Section 4.1*., I faced head on the challenge of providing a positive account of the HOP. While the theoretical frameworks laid down by phenomenal overflow arguments seemed most promising at a first glance, it was concluded that, not only do such models implicitly entail a dissociation of P-consciousness, but they create more problems than solutions by potentially violating such basic assumptions as that expressed by the consistency thesis (see *Section 3*.).

As counter-intuitive as the notion of "phenomenal zombies" might seem, *Section 4.2*. started out with an optimistic tone, proposing that perhaps the HOP might be an analogue of "super-blindsight". It soon had to be conceded, though, that in the case of the HOP, if one were to assume zombiehood, that road would have to be taken all the way. Though there seem to be good reasons to resist the zombie model, it is notoriously hard to provide concise arguments against it without begging the question on the function(s) and/or necessary presence of P-consciousness. Admittedly, a great advantage of the zombie model is that it is both in line with our intuitions concerning the unity of consciousness, and with most of the empirical evidence and theoretical models provided by psychology and the cognitive sciences. The model has one serious problem, though – it assumes that we might "zombily" engage in a relatively sophisticated discussion on our "unconscious" pains.

There is no doubt that any model which claims to provide a satisfactory account of the HOP without violating our most central intuitions about the unity of consciousness or the phenomenal character of pain is worthwhile examining. Thus, in *Section 5*., I inquired whether we should ready ourselves for abandoning the simultaneity claim of the HOP, and adhere to the switching model. Though the model has the apparent virtue of accounting for

how one might have a "hyper-illusion" of two co-existing streams of consciousness, as well as the elegance of subsuming the case of the HOP under the general phenomena of hypnotic amnesia, it seems empirically unfounded, in addition to which it seems to have dire conceptual and/or theoretical consequences. Thus, while this model respects the unity of consciousness, it only does so at the price of abandoning such central intuitions as the continuity of consciousness and the persistence of subjects through time.

My tentative conclusion, then, is that we are ultimately faced with three bad choices: the zombie model, the switching model and the dissociation model. The zombie model is the most attractive in the sense that it has no added complications beyond the evident – a strict revision of our concepts of pain, voluntary behavior, executive functioning and the like. The switching model is the most appealing in the sense that it does justice both to our conceptions of phenomenality and to our conceptions of the unity of consciousness. Yet due to its potentially radical implications concerning temporal continuity and the persistence of the self, this model might be the least desirable from a metaphysical point of view. If the zombie model and the switching model are to be rejected, then I believe we have no more choices – we have to take a deep breath, and concede to the empirical possibility of the dissociation of consciousness.

I myself am as ambivalent as ever as to which model to opt for. Of course, I am not the first to struggle with finding an adequate framework for the HOP: more than a century ago, after his own exposition of this topic, himself being unable to decide, William James noted that, "On the whole it is best to abstain from conclusion" (1890/1983, p. 210). Ironically, he also believed that "The science of the near future will doubtless answer this question more wisely than we can now" (ibid.). If the near future didn't, perhaps the far future will – or perhaps neither will. Either way, I hope the stakes are now clear.

References

Alter, T. (2008). Phenomenal knowledge without experience. In E. Wright (ed.), *The Case for Qualia*. Cambridge, MA: MIT Press, pp. 247–267.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (ed.), *Groups, Leadership and Men: Research in Human Relations*. Oxford: Carnegie Press, pp. 177–190.

Aydede, M. (2010). "Pain". In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), URL = <http://plato.stanford.edu/archives/spr2010/entries/pain/>.

Baars, B. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, *6* (1), 47–52.

Bayne, T. (2007). Hypnosis and the unity of consciousness. In G. A. Jamieson (ed.), *Hypnosis and Conscious States: The Cognitive Neuroscience Perspective*. New York: Oxford University Press, pp. 93–109.

Bayne, T. (2008). The unity of consciousness and the split-brain syndrome. *The Journal of Philosophy*, *105* (6), 277–300.

Bayne, T. (2009). Unity of consciousness. *Scholarpedia, 4* (2), 7414 (revision #52128). URL = http://www.scholarpedia.org/article/Unity_of_consciousness

Bayne, T. and Chalmers, D. (2003) What is the unity of consciousness?. URL = <http://consc.net/papers/unity.pdf> (last accessed: May 16, 2010). Originally appeared in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford: Oxford University Press, pp. 23–58.

Binet, A. (1890). *On Double Consciousness: Experimental Psychological Studies*. Chicago: The Open Court Publishing Company, URL = <http://www.archive.org/stream/ondoubleconsciou00bine#page/n0/mode/2up>.

Block, N. (1995/1997). On a confusion about a function of consciousness. In N. Block, O. Flanagan and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates. Cambridge*, MA: MIT Press, pp. 375–417.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30* (4), 481–548.

Block, N. (2008). Consciousness and cognitive access. *Proceedings of the Aristotelian Society, CVIII* (part 3), 289–317.

Bowers, K. S. (1994). Dissociated control, imagination, and the phenomenology of dissociation. In D. Spiegel (ed.), D*issociation: Culture, Mind, and Body*. Washington, DC, USA: American Psychiatric Press, pp. 21–38.

Bowers, K. S. and Davidson, T. M. (1991). A neodissociative critique of spanos's social-psychological model of hypnosis. In S. J. Lynn and J. W. Rhue (eds.), Theories of Hypnosis: Current Models and Perspectives. New York: The Guilford Press, pp. 105–143.

Breuer, J. and Freud, S. (1895/2007). The psychic mechanism of hysterical phenomena. In J. Breuer (ed.), *Studies in Hysteria*. Boston: Beacon Press, pp. 1–132. (transl. A. A. Brill)

Brook, A. and Raymont, P. (2009). "The Unity of Consciousness". In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), URL = <http://plato.stanford.edu/archives/win2009/entries/conscio usness-unity/>.

Csíkszentmihályi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.

Cosmelli, D. and Thompson, E. (2007). Mountains and valleys: Binocular rivalry and the flow of experience. *Consciousness and Cognition, 16* (3), 623–641.

Dennett, D. C. (1978). Why can't you make a computer that feels pain. *Synthese, 38* (3), 415–456.

De Pascalis, V. (2007). Phase-ordered gamma oscillations and the modulation of hypnotic experience. In G. A. Jamieson (ed.), *Hypnosis and Conscious States: The Cognitive Neuroscience Perspective*. New York: Oxford University Press, pp. 67–89.

Dienes, Z. and Perner, J. (2007). Executive control without conscious awareness: The cold control theory of hypnosis. In G. A. Jamieson (ed.), *Hypnosis and Conscious States: The Cognitive Neuroscience Perspective*. New York: Oxford University Press, pp. 293–314.

Faymonville, M-E., Roediger, L., Fiore, G. D., Delgueldre, C., Phillips, C., Lamy, M., Luxen, A., Maquet, P. and Laureys, S. (2003). Increased cerebral functional connectivity underlying the antinociceptive effects of hypnosis. *Cognitive Brain Research, 17* (2), 255–262.

Glaser, J. and Kihlstrom, J. F. (2005). Compensatory automaticity: Unconscious volition is not an oxymoron. In R. R. Hassin, J. S. Uleman, and J. A. Bargh (eds.), *The New Unconscious*. New York: Oxford University Press, pp. 171–195.

Goldstein, A. and Hilgard, E. R. (1975). Failure of the opiate antagonist naloxone to modify hypnotic analgesia. *Proceedings of the National Academy of Sciences*, USA, *72* (6), 2042–2043.

Grahek, N. (2007). *Feeling pain and being in pain* (second ed.). Cambridge, MA: MIT Press.

Hargadon, R., Bowers, K. S. and Woody, E. Z. (1995). Does counterpain imagery mediate hypnotic analgesia? *Journal of Abnormal Psychology, 104* (3), 508–516.

Harnad, S. (2001). Explaining the mind: Problems, problems. *The Sciences, 41* (2), 36–43.

Hassin, R. R. (2005). Nonconscious control and implicit working memory. In J. R. Hassin, J. S. Uleman, and J. A. Bargh (eds.), *The New Unconscious*. New York: Oxford University Press, pp. 196–224.

Hilgard, E. R. (1973). A neodissociation interpretation of pain reduction in hypnosis. *Psychological Review, 80* (5), 396–411.

Hilgard, E. R. (1977). *Divided Consciousness: Multiple Controls in Human Thought and Action* (expanded ed.). New York: Wiley-Interscience.

Hilgard, E. R. and Hilgard, J. R. (1994). *Hypnosis in the Relief of Pain*. Levittown, PA: Brunner/Mazel Publishers.

Hilgard, E. R., Hilgard, J. R., Macdonald, H., Morgan, A. H. and Johnson, L. S. (1978). Covert pain in hypnotic analgesia: Its reality as tested by the real-simulator design. *Journal of Abnormal Psychology, 87* (6), 655–63.

Hilgard, E. R., Morgan, A. H., and Macdonald, H. (1975). Pain and dissociation in the cold pressor test: A study of hypnotic analgesia with "hidden reports" through automatic key pressing and automatic talking. *Journal of Abnormal Psychology, 84* (3), 280–289.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly, 32* (127), 127–136.

James, W. (1890/1983). *The Principles of Psychology*. Cambridge, MA – London: Harvard University Press.

Kallio, S. and Revonsuo, A. (2003). Hypnotic phenomena and altered states of consciousness: A multilevel framework of description and explanation. *Contemporary Hypnosis, 20* (3), 111–164.

Kallio, S. and Revonsuo, A. (2005). The observer remains hidden. *Contemporary Hypnosis, 22* (3), 138–143.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237* (4821), 1445–1452.

Kihlstrom, J. F. (1998). Dissociations and dissociation theory in hypnosis: Comment on Kirsch and Lynn (1998). *Psychological Bulletin, 123* (2), 186–191.

Kihlstrom, J. F. (2005). Dissociative disorders. *Annual Review of Clinical Psychology, 1*, 227–253.

Kihlstrom, J. F. (in press). Prospects for de-automatization [commentary]. *Consciousness and Cognition*. doi: 10.1016/j.concog.2010.03.004

Kihlstrom, J. F., Barnhardt, T. M. and Tataryn, D. J. (1992). Implicit perception. In R.F. Bornstein and T.S. Pittman (eds.), *Perception without Awareness*. New York: Guilford, pp. 17–54.

Kihstrom, J. F. and Barnier, A. J. (2005). The hidden observer: A straw horse, undeservedly flogged. *Contemporary Hypnosis, 22* (3), 144–151.

Kihlstrom, J. F. and McConkey, K. M. (1990). William James and Hypnosis: A Centennial Reflection. *Psychological Science, 1*(3), 174–178.

Kihlstrom, J. F., Mulvaney, S., Tobias, B. A. and Tobis, I. P. (2000). The emotional unconscious. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas, and P. M. Niedenthal (eds.), *Counterpoints: Cognition and Emotion*. New York: Oxford University Press, pp. 30–86.

Kinnunen, T., Zamansky, H. S. and Block, M. L. (1994). Is the hypnotized subject lying? *Journal of Abnormal Psychology, 103* (2), 184–191.

Kirsch, I. and Lynn, S. J. (1998). Dissociation theories in hypnosis. *Psychological Bulletin, 123* (1), 100–115.

Knox, V. J., Morgan, A. H., and Hilgard, E. R. (1974). Pain and suffering in ischemia: The paradox of hypnotically suggested anesthesia as contradicted by the reports from the "hidden observer". *Archives of General Psychiatry, 30* (6), 840–847.

LaBerge, S. (1990). Lucid dreaming: Psychophysiological studies of consciousness during REM sleep. In R. R. Bootzen, J. F. Kihlstrom and D. L. Schacter (eds.), Sleep and Cognition. Washington, D.C.: American Psychological Association, pp. 109–126.

Levine, J. (2007). Two kinds of access [commentary]. *Behavioral and Brain Sciences, 30* (5/6), 514–515.

Loeser, J. D. and Melzack, R. (1999). Pain: An overview. *Lancet, 353* (9164), 1607–1609.

Lynn, S. J. and Rhue, J. W. (1991). Hypnosis theories: Themes, variations, and research directions. In S. J. Lynn and J. W. Rhue (eds.), *Theories of Hypnosis: Current Models and Perspectives*. New York: The Guilford Press, pp. 601–626.

Maddula, M., Lutton, S. and Keegan, B. (2009). Anton's syndrome due to cerebrovascular disease: A case report. *Journal of Medical Case Reports, 3* (1), 9028.

Montgomery, G. H., DuHamel, K. N. and Redd, W. H. (2000). A meta-analysis of hypnotically induced analgesia: How effective is hypnosis? *International Journal of Clinical and Experimental Hypnosis, 48* (2), 138–153.

Melzack, R. (1993). Pain: Past, present and future. *Canadian Journal of Experimental Psychology, 47* (4), 615–629.

Melzack, R. and Casey, K. L. (1968). Sensory, motivational, and central control determinants of pain*:* A new conceptual model. In D. Kenshalo (ed.), *The Skin Senses*. Springfield, IL: Chas T. Thomas, pp. 423–443.

Miller, M. E. and Bowers, K. S. (1993). Hypnotic analgesia: Dissociated experience or dissociated control?. *Journal of Abnormal Psychology, 102* (1), 29–38.

Miller, M. F., Barabasz, A. F. and Barabasz, M. (1991). Effects of active alert and relaxation hypnotic inductions on cold pressor pain. *Journal of Abnormal Psychology, 100* (2), 223–226.

Montgomery, G. H., Duhamel, K. N. and Redd, W. H. (2000). A meta-analysis of hypnotically induced analgesia: How effective is hypnosis? *International Journal of Clinical and Experimental Hypnosis, 48* (2), 138–153.

Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese, 22* (3/4), 396–413.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83* (4), 435–50.

Orne, M.T. (1959). The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology, 58* (3), 277–299.

Puccetti, R. (1981). The case for mental duality: Evidence from split-brain data and other considerations. *Behavioral and Brain Sciences, 4* (1), 93–123.

Rainville, P., Carrier. B., Hofbauer, R. K., Bushnell, M. C. and Duncan, G. H. (1999). Dissociation of sensory and affective dimensions of pain using hypnotic modulation. *Pain, 82* (2), 159–171.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., and Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science, 277* (5328), 968–971.

Raz, A., Shapiro, T., Fan, J., and Posner, M. I. (2002). Hypnotic suggestion and the modulation of Stroop interference. *Archives of General Psychiatry, 59* (12), 1155–1161.

Robinson, H. (1989). A dualist account of embodiment. In J. R. Smythies and J. Beloff (eds.), *The Case for Dualism.* Charlottesville: University Press of Virginia, pp. 43–58.

Schacter, D. L. (2000). Understanding implicit memory: A cognitive neuroscience approach. In M. S. Gazzaniga (ed.), *Cognitive Neuroscience: A Reader*. Oxford: Blackewell Publishers, pp. 305-324.

Shallice, T. (1997). Modularity and consciousness. In N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: The MIT Press, pp. 255–276.

Spanos, N. P. (1986). Hypnotic behavior: A social-psychological interpretation of amnesia, analgesia, and "trance logic". *Behavioral and Brain Sciences, 9* (3), 449–502.

Spanos, N. P. and Hewitt, E. C. (1980). The hidden observer in hypnotic analgesia: Discovery or experimental creation? *Journal of Personality and Social Psychology, 39* (6), 1201–1214.

Sperling, G. (1960). The information available in brief visual representations. *Psychological Monographs: General and Applied, 74* (11), 1–29.

Sperry, R. W., Gazzaniga, M. S. and Bogen, J. E. (1969). Interhemispheric relationships. the neocortical commissures; syndromes of hemispheric disconnection. In P. J. Vinken and G. W. Bruyn (eds.), *Handbook of Clinical Neurology, Vol. 4.* Amsterdam: North-Holland Publishing Company.

Spiegel, D. (1990). Hypnosis, dissociation, and trauma: Hidden and overt observers. In J. L. Singer (ed.), *Repression and Dissociation: Implications for Personality Theory, Psychopathology, and Health.*. Chicago: The University of Chicago Press, pp. 121–142.

Tononi, G., Srinivasan, R., Russell, D. P., and Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Sciences of the United States of America, 95* (6), 3198–3203.

Weinberger, J. (2000). William James and the unconscious: Redressing a century-old misunderstanding. *Psychological Science, 11* (6), 439–445.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford: Oxford University Press.

Zamansky, H. S. and Bartis, S. P. (1985). The dissociation of an experience: The hidden observer observed. *Journal of Abnormal Psychology*, 94 (*3*), 243–248.