frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 27 September 2011
doi: 10.3389/fpsyg.2011.00224

Have we vindicated the motivational unconscious yet? A conceptual review

Alexandre Billon*

Department of Philosophy, Université Lille-Nord de France, UMR 8163, Lille, France

Edited by:
Thierry Simonelli, Independent practice, Luxembourg

Reviewed by:
Mathieu Arminjon, Hôpitaux Universitaires de Genève, Switzerland
Gilles Arnaud, University of Toulouse, France

Correspondence:
Alexandre Billon, Université de Lille 3, Bureau B4.147 B.P. 60149, Villeneuve d'Ascq CEDEX, Lille, France.
e-mail: abillon@gmail.com

Motivationally unconscious (M-unconscious) states are unconscious states that can directly motivate a subject's behavior and whose unconscious character typically results from a form of repression. The basic argument for M-unconscious states claims that they provide the best explanation for some seemingly non-rational behaviors, like akrasia, impulsivity, or apparent self-deception. This basic argument has been challenged on theoretical, empirical, and conceptual grounds. Drawing on recent works on apparent self-deception and on the "cognitive unconscious" I assess those objections. I argue that (i) even if there is a good theoretical argument for its existence, (ii) most empirical vindications of the M-unconscious miss their target. (iii) As for the conceptual objections, they compel us to modify the classical picture of the M-unconscious. I conclude that M-unconscious states and processes must be *affective* states and processes that the subject really *feels and experiences* – and which are in this sense conscious – even though they are not, or not well, *cognitively accessible* to him. Dual-process psychology and the literature on cold–hot empathy gaps partly support the existence of such M-unconscious states.

Keywords: unconscious, self-deception, consciousness, confabulation, repression, motivation, dual process psychology, dissociations

It is usually easy to explain someone's behavior. We can often read the motives that explain his actions directly on his moves and expressions. When we cannot, it is always possible to ask the person the reason for his actions. I would normally explain why John pulled the trigger of the gun while pointing it at his wife by citing his desire to get rid of her and his belief that this is an excellent way of doing just that. If I had some doubts, I could ask John. He might tell me that he's no murderer and he just wanted to wound her. I might accordingly revise my explanation. Provided that John is "in good faith," I should end up with a pretty good explanation. This explanation will have the following form: a motive for an action A – roughly[1] a desire for a given outcome and the belief that A is the best way to achieve this outcome – explains the action A. There are much tougher cases though, in which this method will not be of much help, even if the subject whose behavior we are trying to interpret can answer all our questions, and even if he is perfectly honest.

AKRASIA AND IMPULSIVITY

The cases I am thinking of have bewildered philosophers ever since Plato. They involve actions that do not seem rational. Consider for example, weakness of the will or *akrasia*. Someone who is akratic seems to be acting against his preferences. A typical example involves someone, say Paul, who sees someone else's wallet on the ground, who knows he should not take it, wants to be honest "and everything", prefers, all things considered not to take it, . . . but yet, steals the wallet. Paul "sees the better and approves it, but he follows the worse". Notice that Paul might, but need not be acting impulsively. He might as well pursue the worse, so to speak, "in cold blood". In any case, he will in good faith deny having the preference that would explain his behavior, so it seems reasonable not to attribute him this preference. But how then should we explain his misdemeanor? (compare: "John chose A over B even though he judged A to be more expensive/less pretty. . ." to "John chose A over B even though he judged that all things considered, it was better to chose B").

APPARENT SELF-DECEPTION

Consider also cases of "self-deception". A victim of self-deception seems to be deceiving himself into believing something false. He believes something against the evidence, and he seems to do so intentionally. For example, Mary, who was recently diagnosed with a cancer claims that the diagnosis is misguided and she really seems to believe it, but she avoids her physicians as well as conversations on medical topics. She seems somehow to be lying to herself about her disease. Self-deception names both the phenomenon "someone seems to be deceiving oneself" and its natural explanation "this person actually deceives herself". We shall see that many refuse this natural explanation so it will be useful to have two different names here. We will call apparent self-deception the phenomenon, and self-deception its natural explanation. In any case, being "apparently self-deceived" Mary will in good faith deny having the intention to blind herself, so it will be difficult to explain why she does not acknowledge her disease and why she avoids doctors.

[1]See (Pacherie, 2008a) for some problems with this basic account of motives and for a sophisticated alternative.

THE MOTIVATIONAL UNCONSCIOUS (M-UNCONSCIOUS)

Because the motives that would most naturally explain the mysterious actions of Mary and Paul are not acknowledged by them, it might be tempting, in both cases, to claim that they have *unconscious motives*. There is an additional, and so to speak logical, reason to claim that there must be unconscious motives in the case of apparent self-deception. If Mary really deceived herself into falsely believing that she is perfectly healthy, as it seems she did, it would be paradoxical to suppose that her motive – her belief that she is sick and her desire to deceive herself into believing that she is not – was conscious. We cannot indeed be deceived by someone when we are aware of his or her desires to deceive us, or even when we are aware that what he or she says is false. The only way to solve this paradox, it might be claimed, and to make sense of the possibility of deceiving oneself, is to posit unconscious motives. If Mary only unconsciously knows that she is ill, and if she only unconsciously desires to believe that she is not ill, then she might actually deceive herself. Notice that this last argument does not rely on the claim that Mary is in good faith when she speaks about her motives and that conscious motives are always available to introspection. It only relies on the claim that she is really deceiving herself, as she seems to be, and that this logically implies unconscious motives. On that account, apparent self-deception would result from self-deception, and self-deception would not only prevent the subject of being conscious of some things: it would manage in doing so because it is motivated by unconscious beliefs and desires. Apparent self-deception, that is, would be a form of *repression*.

We can call motivational unconscious (M-unconscious) the kind of unconscious posited by those explanations. The M-unconscious is defined by the three following constraints:

- I. It is a form of unconscious: M-unconscious states are unconscious.
- II. It is a motivating unconscious: M-unconscious states can be motives for actions.
- III. It is a motivated unconscious: the unconscious character of M-unconscious states typically results from a form of repression.

THE BASIC THEORETICAL ARGUMENT FOR THE M-UNCONSCIOUS

The *basic argument* for the M-unconscious is the one I just gave. It argues for the existence of M-unconscious states on the grounds that they provide the best explanation of some seemingly non-rational behaviors like self-deception. This is, roughly, Freud's first argument for the unconscious in his 1915 paper *The unconscious* (Freud, 1974, p. 2991–3025). Actually, even before he used the notion of unconscious, Freud started to interpret phobias, obsessional compulsive disorders, and conversion disorders in terms of self-deception, as "neuro-psychoses of defense" Freud (1974, pp. 301–313). He then explained self-deception in terms of unconscious mental states and argued that M-unconscious states also manifest themselves in other, benign, phenomena such as dreams and lapsus.

It should be noted that even though Freud is probably one of the greatest theorists of such a M-unconscious, he is neither the first, nor the last one. Nussbaum (1994, p. 133) convincingly argues that Lucretius, and probably Epicurus, appealed to an M-unconscious fear of death in order to explain some of our irrational behaviors. In the same vein, some contemporary psychologists have also postulated M-unconscious states quite unlike those referred to by Freud (Yalom, 1980 is an example). The notion of M-unconscious is thus wider than that of the "Freudian unconscious".

The recent fate of this notion, however, was shaped by that of Freud's legacy. The M-unconscious endured attacks on at least four partially distinct fronts in the last century.

THEORETICAL OBJECTIONS

On the theoretical side, it was argued that the basic argument for the M-unconscious is not sound. It is an abductive argument: an argument "to the best explanation". It argues for the existence of M-unconscious states on the ground that they provide the best explanation of some of our behaviors. In order for such an argument to be conclusive though, it is not enough to claim that there are real cases of akrasia and apparent self-deception. It must be shown that the rival explanations of the same phenomena are not better. As we shall see, Jaspers (1997) already disputed that claim, and many philosophers have recently put forward very neat accounts of apparent self-deception which do not appeal to unconscious states.

EMPIRICAL (AND EPISTEMOLOGICAL) OBJECTIONS

The basic argument is also a *theoretical* argument: it relies on folk-psychological observations, clinical reports, and individual narratives rather than on controlled empirical experiments. It is moreover dubious that M-unconscious states have been directly observed. Some have argued, in the wake of logical positivists, that the reality of apparent self-deception and of the M-unconscious should not accordingly be granted. It has even been argued that the M-unconscious had not only failed to be empirically ascertained, but also that *it had to*, for principled reasons. I will not, however, consider those well known epistemological objections here.

CONCEPTUAL OBJECTIONS

Finally, it has been claimed that the very concept of M-unconscious is inconsistent or at least very problematic and that it should be reformed, if not eliminated (Sartre, 1976; Wittgenstein, 1982).

OUTLINE

Those attacks probably contributed to the withdrawal of the M-unconscious from scientific psychology. Studies on the M-unconscious were substituted by experiments focused on the "cognitive unconscious" which is not defined by its connection to motivation but only by its unconscious character. The latter seemed much simpler – one might say much dumber – and for that reason much less problematic than the M-unconscious.

The face of the cognitive unconscious has however drastically changed since Kihlstrom (1987) contrasted it with the "psychoanalytic unconscious". Recent developments in cognitive neurosciences have shown that unconscious states are many, and that they can occupy a wide array of functions long thought to be the prerogative of consciousness. Not only can we perceive objects unconsciously, we can also perceive some of the high level features

of those objects, like the meaning of words (Dehaene and Naccache, 2001). We can learn rules unconsciously (Cleeramns et al., 1998). We can unconsciously appraise things or persons (Fazio and Williams, 1986; Bargh et al., 1996). Some results even suggest that self-regulation and metacognition can also occur unconsciously (Hassin et al., 2005). In the meantime, philosophers of cognitive science have shown that we could meet some of the difficulties raised by the idea of a rich unconscious mental life, of unconscious representations and so on. All this led Uleman (2005), in the introduction of a collection of essays devoted to the topic, to conclude that we should now talk of the "new unconscious" and that "the list of psychological processes carried out in the new unconscious is so extensive that it raises two questions: what, if anything, cannot be done without awareness? What is consciousness for?".

By an interesting twist of fate, this new understanding of the cognitive unconscious has also led to an important revival of some Freudian ideas, some researchers even claiming that after years of neglect, the Freudian unconscious had finally been vindicated.

The aim of this paper is to take advantage of those recent developments to reassess the basic argument for a motivational unconscious at a theoretical, empirical, and conceptual level. This task is somehow too wide and I will have to restrict it by focusing on the version of the basic argument that appeals to apparent self-deception. This limitation is justified by the central role of apparent self-deception in the motivational unconscious: repression is nothing but the way partisans of the motivational unconscious explain apparent self-deception. It is also justified by the quality and the variety of the researches that have been devoted on this topic these last 30 years. I will first show that the basic argument can be rescued and developed into a cogent *general theoretical argument* for the M-unconscious. This general argument concludes that the M-unconscious provides the best explanation of apparent self-deception, or rather of what we know of it from folk-psychological observations, psychiatric case histories, and first-person narratives (see Section 2).

I will then investigate some recent "empirical vindications" of the motivational unconscious (see Section 3). Here again, I will have to restrict my overview to three influential series of studies that try to empirically certify the existence of apparent self-deception in order to prove the existence of the M-unconscious. Despite its limitation, this overview will allow us to spot some shortcomings shared by many studies of the same vein, shortcomings which explain what I take to be the rather limited success of this "vindicating enterprise."

Finally I will tackle what I perceive to be the two major conceptual problems posed by the M-unconscious (see Section 4).

The outcome of this threefold review will be an original picture of the M-unconscious which I call affective and neo-dissociationist. According to that picture, M-unconscious states and processes are affective states and processes that the subject really feels and experiences – and which are in this sense conscious – even though they are not, or not well, cognitively accessible to him. This picture of the M-unconscious is affective in that M-unconscious states and processes are essentially affective. It is neo-dissociationist in that, even if M-unconscious states are in some sense unconscious, they are in another important sense conscious, the role of repression being to maintain the dissociation of two forms of consciousness rather than to expel some states from a monolithic consciousness. Dual-process psychology and the literature on cold–hot empathy gaps partly support this version of the M-unconscious. I argue that the latter does not only have some empirical plausibility. It is also the only one that can escape the two conceptual problems (see Section 5).

Those four sections are not unrelated. Indeed we will see that the alternative explanations of apparent self-deception found in the first part will help in criticizing the empirical studies in the second part, and that the conceptual problems which afflict the M-unconscious will allow us to diagnose the relative failure of the empirical vindication. I will indeed argue that the M-unconscious must be quite unlike what we have been looking for in most empirical studies.

1. PERSONAL LEVEL EXPLANATIONS AND THE M-UNCONSCIOUS

Before addressing the theoretical, the empirical, and the conceptual objections confronting the M-unconscious we need to introduce some terminology and to make the definition of the M-unconscious more articulate.

We shall see, in the next section, that there are very simple and now very common explanations that could replace the ones in terms of M-unconscious. The explanations I am thinking of are framed in terms of mere "unconscious" brain states, activation levels, and neurotransmitters. They are what we might call *purely* neurophysiological explanations. Such neurophysiological explanations have been put forward to explain depression and addiction[2], and there is no reason to suppose that similar explanations are not available in the case of apparent self-deception. It is important to understand that explanations of this kind are radically different from the ones in terms of the M-unconscious.

1.1. MOTIVES AND CAUSES

The distinction between *motives* and *mere causes* is in this respect crucial. By citing motives, folk-psychological explanations do not only allow to predict actions, they also "make sense" of them. They justify them. If John desires to wound his wife and points his gun at her, it makes sense for him to pull the trigger. His action is in that respect justified by his motive. By only citing causes, purely neurophysiological accounts also allow to predict actions, but they do not make sense of them[3]. When a judge asks you "why did you do it?", he is asking for motives, not for mere causes. When a neurophysiologist asks the same question, he might be asking for mere causes. It is in this technical sense of motivation that M-unconscious states can be motives for action. By positing some states which are unconscious but are motives nonetheless, explanations in terms of M-unconscious accordingly prolong rather than revise folk-psychological explanations.

[2] See Section 5.2.2.

[3] It should be emphasized that an event can causally influence the motives of a subject without *ipso facto* constituting a motive of the subject. The studies of Pessiglione et al. (2007), Schmidt et al. (2010) for example show that some unconscious causes influence our motivation, not that our motivation, in the technical sense introduced here, can be unconscious.

1.2. PERSONAL AND SUBPERSONAL EXPLANATIONS

It is customary, since Dennett (1969), to say that motives and the kind of "sense-making" explanations in which they figure belong to the *personal level* as opposed to the *subpersonal* or even *parapersonal* level. This terminology reflects the fact that this kind of explanation mentions states, events and processes that we would naturally ascribe to a person. Other explanations, on the other hand, can mention states, events or processes that we would naturally attribute to a proper part ("sub") of a person like his brain, or to quasi-subjects ("para") within the person. Neuropsychological explanations are, for example, typically subpersonal.

1.3. THE THREE CONSTRAINTS ON A MOTIVATIONAL UNCONSCIOUS

These distinctions can help us make the three constraints on a M-unconscious more precise. M-unconscious states (processes, events. . .) are not only unconscious (I). They are also *motivating*:

II. The M-unconscious is a motivating unconscious. M-unconscious states can be motives of the subject's actions, they are states that can explain those actions at a personal level.

M-unconscious states are also typically *motivated*: their unconscious nature results typically results from repression. Repression in this sense includes the expulsion from consciousness of a conscious state as well as the "defense" processes that maintain an unconscious state away from consciousness. Here again, it should be insisted that repression is a personal level phenomenon that can be compared to hiding something from someone. It is something the subject does for a motive, hence something he does intentionally[4]. It is not a subpersonal happening, and it is not the action of a small homunculus within the subject. As it is intentional, it is not the mere *collateral effect* of some intentional actions of the subject either. When the bombing of a bridge causes civilian losses, the latter are sometimes unintended. They are a mere collateral effect of an intentional action. This implies that "killing civilians" is not an intentional action but "bombing the bridge" is, even though both expressions can be used to describe what the subject did. Repressing a mental state is in that respect like bombing the bridge rather than like killing civilians:

III. The M-unconscious is a motivated unconscious. M-unconscious states typically result from a form of repression, which is an intentional action of the subject[5].

[4]By that I do not mean that it is a cautiously deliberated and reflected on action. Passion crimes are intentional in my sense. Many authors have insisted that Freud did not really mean, or at least not always, that repression is intentional (Laplanche and Pontalis, 1968, pp. 394–395; Schacter, 2001; Gardner, 2006). I take it, however, that they have a much more substantial understanding of "intentional", under which it implies a form of rationality and strategic planning (this is clear for Gardner, 2006 and Laplanche and Pontalis, 1968). Under my acceptation, which is widespread in philosophical community, intentional, and motivated are synonymous. Any action, be it impulsive, is intentional under some description.

[5]Even though Freud might have thought, like e.g., Melanie Klein, that some unconscious states cannot be conscious at all, and thus cannot result from repression, the second constraint is clearly endorsed by Freud. He says in 1923 that "we obtain our concept of the unconscious from the theory of repression. The repressed is the prototype of the unconscious for us (Freud, 1974, p. 3949)."

2. THEORETICAL OBJECTIONS AGAINST A M-UNCONSCIOUS

The basic argument for a M-unconscious has been challenged, as we said, on theoretical grounds. I will try to show that the M-unconscious can actually withstand those criticisms quite well.

2.1. AGAINST SUBPERSONAL ACCOUNTS OF APPARENT SELF-DECEPTION

The simplest rivals to the explanation of apparent self-deception in terms of M-unconscious states are framed at a purely subpersonal level. Such explanations would provide unconscious *causes* for the behaviors under scrutiny, and they would not commit to the existence of unconscious *motives*. Correspondingly, they would not try to make sense of those behaviors. The gist of Jaspers (1997)'s early objections to Freud was precisely that the latter had not shown the need to reject such subpersonal explanations of those non-rational behaviors in favor of personal level ones[6].

Why should we favor personal level explanations? One answer has it that personal level explanations should, when they are possible, always be favored to subpersonal explanations. It could be claimed, for example, that we should by default maximize the intelligibility of the subjects we interpret (this is a – rather hasty – way to understand the "principle of charity"; Davidson, 1984). In such a bold form however, this claim is not tenable. I just remembered a minute ago that I needed to buy some bread. Why a minute ago rather than a minute earlier? This question has an answer in subpersonal term. I take it however that it would in general be totally unreasonable to look for another, complementary, answer framed in personal terms. At least in normal circumstances, there is just no plausible explanation of such a phenomenon in personal terms[7].

What is true is that in psychology, like everywhere, a good explanation should by default *save the appearances* and that some of the phenomena we are trying to explain, unlike remembering a minute ago to buy some bread, do indeed seem to be phenomena that can be explained in personal terms. By definition, apparently self-deceived subjects seem to be deceiving themselves. If we want to save this appearance we will have to explain this phenomenon in personal terms. Moreover, and pace eliminativism (Churchland, 1981), personal level explanations have many virtues, which might justify that we should favor them when they are not too implausible. They have some explanatory virtues: they are simple and easy to understand, they are also quite general and have relatively good predictive capacities. Although it is probably more controversial, I take it that they also have some moral virtues. They make it easier, for example, to consider the subject as responsible of his deeds and to empathize with him.

2.2. THE PARADOXES OF SELF-DECEPTION

Those are reasons to favor, by default at least, a personal level explanation of apparent self-deception. Now personal level explanations can appeal to conscious states as well as unconscious states. Why

[6]Jaspers (1997) characterizes personal level explanations by the "genetic understanding" they provide and he opposes them to mere causal explanations.

[7]I owe this example to Tim Bayne.

should we prefer unconscious states in some cases? Why should we, furthermore, appeal to repression? We outlined a *basic argument* to the effect that personal level explanations of apparent self-deception must invoke unconscious motives and repression. This argument claims that other explanations of the same phenomenon will lead to paradoxes. Even if their primary target was not usually the M-unconscious, in the last 40 years, philosophers of mind have precisely articulated those paradoxes and have shown that there are many other ways to escape them and to explain apparent self-deception which do not appeal to unconscious states. Those alternative solutions to the paradoxes, and the alternative explanations of apparent self-deception which rely on them, threaten the basic argument. We should accordingly review those articulate versions of the paradoxes and assess their various solutions.

Self-deception can give rise to two paradoxes (Mele, 2001; Davidson, 2004).

The static paradox of self-deception

1. In order to deceive someone into believing P, as opposed to merely induce someone into believing P, one has to believe the contrary proposition ∼ P.
2. In order to be deceived into believing P one must believe P.
3. If someone believes P and believes ∼ P he believes (P and ∼ P).
4. So by (1–3), in order to deceive oneself into believing P, one must believe (P and ∼ P).

This is paradoxical because although apparent self-deception is certainly irrational it does not seem so irrational as to involve the belief in a bold contradiction.

The strategic paradox has to do with the absurdity of the *project* or intention of deceiving myself rather than with the absurdity of the *state* of being self-deceived.

The strategic paradox of self-deception

1. In order to deceive someone intentionally I must know (and hence believe) that I am trying to deceive him.
2. In order for someone to be deceived by me, he must believe that I am not trying to deceive him.
3. If I believe that I am trying to deceive myself and if I believe that I am not, I believe that (I am and I am not trying to deceive myself).
4. So by (1–3), in order to deceive myself I must believe that (I am and I am not trying to deceive myself).

This is paradoxical, again, because apparent self-deception does not seem *that* irrational.

The M-unconscious provides, as we said, a very natural solution to those paradoxes. When someone apparently deceives herself into believing P, she *consciously* believes P and *unconsciously* believes ∼ P. This solves the static paradox because one can believe P-consciously and believe not P unconsciously without believing (P and ∼ P). Moreover, the intention of deceiving oneself must be itself unconscious, so the subject may not believe that he has it. This solves the strategic paradox.

According to this explanation, self-deception stems from an unconscious motive (the intention to deceive oneself) and results in a repressed, unconscious belief (the belief that ∼ P). The natural character of this solution does not suffice, however, to establish that it is the best solution to the paradoxes, and that the M-unconscious accordingly provides the best personal level explanation of apparent self-deception. There has been a considerable amount of philosophical and psychological theorizing on apparent self-deception in the last 40 years and many have proposed alternative explanations of this phenomenon that are spelled in personal terms, but which do not appeal to unconscious states. In order to be conclusive, the general argument for the M-unconscious must show that those explanations are not as good as the one in terms of M-unconscious. Reviewing those explanations will prove useful because they provide interesting variants of the Freudian explanations, variants that we should keep in mind when we will assess the empirical evidence for the M-unconscious.

2.3. EXPLAINING APPARENT SELF-DECEPTION IN TERMS OF A DIVIDED SELF

Ironically enough, with one notable exception (Fingarette, 1969), the debate on apparent self-deception has largely ignored the latter Freudian solution. It has concentrated instead on a close cousin of it, which postulates, instead, a separation within the self between two centers of beliefs and agency.

2.3.1. Cognitive separations and the static paradox

One can indeed solve the static paradox by postulating that the conflicting beliefs are both conscious but that they are hidden to each other by a form of separation. This involves construing the self as divided into different centers of conscious beliefs.

2.3.2. Agentive separations and the strategic paradox

In order to get rid of the strategic paradox we must go further and suppose that the intention to deceive, just like the belief that P which contributes to motivate it, is also separated from the rest of the subject. The separation will thus be *agentive* as well as cognitive.

When a subject is self-deceived, a part of himself would be deceiving another part of himself. Partisans of such a solution disagree on the degree of autonomy one should grant to those parts of the self. Rorty (1988) pictures them as full-blown homunculi. Pears (1987) acknowledges different centers of agency but tries not to picture them as full-blown subjects. However, the mere suggestion that the self is *literally* divided in multiple centers of agency is sufficiently awkward already. It is one thing to say that someone has different, and maybe conflicting, faculties, and inclinations, it is quite another to say that it is constituted by different centers of agency! Davidson (1983) acknowledges the problem. He actually takes it, wrongly I believe, to be a problem afflicting the Freudian conception of the mind (Davidson, 2004, pp. 170–171). He tries to eschew it by talking of *semi*-autonomous divisions or *quasi*-independent structures (Davidson, 2004, p. 181) while making clear that "the idea of a quasi-autonomous division is not one that demands a little agent in the division" (Davidson, 2004, p. 181). Davidson later emphasized that "the image [he] wished to invite was not, then, that of two minds each somehow able to act like an independent agent; the image is rather that of a single mind not wholly integrated; a brain suffering from a perhaps temporary self-inflicted lobotomy." I think these remarks point toward the

right direction, which is, we shall see, that of the M-unconscious approach to self-deception, and of its neo-dissociationist variant (see Section 6.1). The problem is that it is not clear that Davidson has the resources to solve the strategic paradox and to endorse one of those approaches while refusing, as he does, to be spell out more precisely the nature of this lack of integration[8] and to resort to the concept of (un-)consciousness[9].

The divided self approach to apparent self-deception is not only metaphysically extravagant. It suffers from a deeper, seldom noted, problem[10]. It cannot, by itself, explain apparent self-deception. Suppose there are two centers of agency in someone, a deceiver and a deceived one. The deceived center of agency has reasons to be deceived: believing something false will alleviate his pains and worries. But what reasons could the deceiving center have to deceive the deceived center? By definition, two distinct centers of agency can have independent motives for acting. Just like your reasons for pulling the trigger are not in general reasons for me to help you do so, the deceived center's reasons for being deceived will not in general be reasons for the deceiver to do the deceiving. The partisan of the divided self approach owes us a story here, a story which he unfortunately never provides.

It should be noted, by the way that this idea of a divided self, an idea for which Sartre (1992) notoriously criticized Freud and which Davidson also attributes to Freud has not been endorsed by him[11]. As Gardner (2006, pp. 148–153) has argued at length, "The ego, id, and superego, as parts of the soul, do war, but they are not each of them warring souls."

2.4. A FIRST DEFLATIONARY ACCOUNT: APPARENT SELF-DECEPTION AS DIACHRONIC SELF-DECEPTION

Many philosophers have insisted that we can avoid the problems posed by the divided self approach without espousing a Freudian approach to apparent self-deception. They have tried to explain this phenomenon in a *deflationary* way, that is, *without positing a present intention to deceive oneself*. Some have proposed that self-deception is actually diachronic. When I decide not to indicate a future appointment in my notebook, hoping that I will eventually forget it, I can manage to deceive myself without paradox because it is so to speak my present self which deceives my future self. All cases of apparent self-deception, argue the partisans of the diachronic self-deception approach, would be like that. A subject would form the intention at time *t*, to deceive himself at a future time *t'*. The problem with this proposal is that at least in some cases of apparent self-deception – let us call them cases of *severe* apparent self-deception, the subject displays a form of (synchronic) inner conflict. This conflict can be witnessed by an equivocal behavior, as in the example of the self-deceived patient who claims not to suffer from cancer but simultaneously avoids his physicians. In that case as in others, the conflict can also surface in abnormal emotional reactions associated with the assertion of P. Just like liars, subjects of severe forms of apparent self-deception can for example show heightened autonomic response when making their claim (Bonanno et al., 1995). Diachronic accounts cannot explain severe apparent self-deception. They do not, accordingly, threaten the M-unconscious approach.

2.5. A SECOND DEFLATIONARY ACCOUNT: APPARENT SELF-DECEPTION AS MOTIVATIONALLY BIASED BELIEFS

Another deflationary account of apparent self-deception consists in explaining away the impression that the subject intends to deceive himself at one point or another. There is ample empirical evidence that our belief formation mechanisms are often biased by our desires. Mele (1997, 2001) has argued that motivationally biased beliefs could explain cases of apparent self-deception. Consider for example a young man who recently lost his father. To alleviate his pains, he tries to remember the good moments he spent with him, so he focuses on situations in which his father was very nice and he comes to forget darker aspects of his personality. He ends up believing falsely that he was a very gentle man. Here the belief is certainly biased by the young man's desires. His belief might even correspond to what he desires to be the case. But it is better described as a case of *wishful thinking* than as a case of self-deception, for the subject never had to form the intention to believe or to forget anything. His desires biased his belief, and thus caused it, but they did not motivate it. The belief was not formed intentionally, to put the same point in another way, it was only a happy collateral effect. Mele (1997, 2001) claims that all cases of apparent self-deception are like that.

Like the preceding one, this approach cannot however account for the inner conflicts which are characteristic of severe cases of apparent self-deception. It also suffers from a specificity problem (Bermùdez, 2000). We have many desires that do not give rise to apparent self-deception. Theories of apparent self-deception which appeal to the subject's intention to deceive himself can readily account for this specificity of apparent self-deception: apparent self-deception would only arise when a subject not only desires P or even desires to believe P, but when he also *intends* to believe P. It is not obvious that the theory of apparent self-deception as motivationally biased beliefs can give a satisfying account of the specificity of apparent self-deception (although see Mele, 2001, chap. 3).

2.6. THE GENERAL ARGUMENT FOR A M-UNCONSCIOUS

Having reviewed alternative approaches to self-deception and their problems (**Table 1**), we are in a position to state *the general theoretical argument* for the M-unconscious.

[8]To the question "How can a person fail to put the inconsistent or incompatible beliefs together? " he answers "It would be a mistake to try to give a detailed answer to this question here. The point is that people can and do sometimes keep closely related but opposed beliefs apart (Davidson, 2004, p. 211)."

[9]He explicitly denies the need to do so in two places. "I see no obvious reason to suppose one of the territories must be closed to consciousness, whatever exactly that means, but it is clear that the agent cannot survey the whole without erasing the boundaries (Davidson, 2004, p. 211)." "Finally, I must mention the [Freudian] claim that many mental phenomena which normally are accessible to consciousness are sometimes neither conscious nor easily accessible to consciousness. The reason I have said nothing about this claim is that I think the relevant objections to unconscious mental states and events are answered by showing that the theory [the theory of apparent self-deception he defends] is acceptable without them (Davidson, 2004, pp. 185–186)."

[10]Fingarette (1983) evokes this problem while discussing divided self interpretations of the Freudian conception of the mind.

[11]"By the distinction between the "id" and the "ego," Freud has cut the psychic whole into two. I am the ego but I am not the id (. . .) By rejecting the conscious unity of the psyche, Freud is obliged to imply everywhere a magic unity (Sartre, 1992, pp. 50–53)."

Table 1 | The various approaches to apparent self-deception.

Category of	Approach	Problems
Other	Subpersonal approach	Does not "save the appearances"
Substantial approaches	Divided self approach	Metaphysically extravagant, explanatorily inadequate (motivation)
	M-unconscious approach (repression)	
	Neo-dissociationist approach	
Deflationary approach	Diachronic self-deception	Explanatory inadequate (tension)
	Motivationally biased belief (wishful thinking)	Explanatorily inadequate (selectivity and tension)

The general theoretical argument claims that the "M-unconscious approach" which explains apparent self-deception in terms of repression is the best. The neo-dissociationist approach is a variant of the M-conscious approach that will be made explicit in Section 6.1.

1. Exclusion of deflationary approaches. Cases of apparent self-deception, which involve selectivity and inner conflict, are better explained in a non-deflationary way.
2. Folk-psychological observation. Apparent self-deception sometimes involve selectivity and inner conflict.
3. General explanatory principle. A case of apparent self-deception which is better explained in a non-deflationary way involves a form of motivational unconscious.

The general explanatory principle rests on the supposition that the subjects are not too irrational [so that they cannot believe plain contradictions like (P and not P)] and on the rejection of the divided self and the subpersonal approaches to apparent self-deception. The exclusion of deflationary approaches rests on the fact that deflationary approaches of apparent self-deception, which do not posit a present intention to deceive oneself, cannot explain the presence of inner conflicts or of selectivity.

3. EMPIRICAL EVIDENCE FOR A M-UNCONSCIOUS

The theoretical reasons to posit a M-unconscious rely on folk-psychological observations and on the general explanatory principle, which depends itself on contentious methodological claims. It would not weigh much in the face of empirical evidence to the effect that there are no M-unconscious states. The general argument is in that sense provisional and it would be nice if its conclusion could be empirically confirmed.

Many recent studies have been said to provide the confirmation needed. I will review three influential series of studies on memory suppression, confabulation, and anosognosia for plegia respectively. I will argue that they mostly miss their target. Only the studies on anosognosia come close to vindicating the M-unconscious, but they tend to favor a neo-dissociationist variant of the M-unconscious approach, according to which the repressed states are not unambiguously conscious.

3.1. MEMORY SUPPRESSION

In a much quoted study, Anderson and Green (2001) have shown that subjects can intentionally suppress unwanted memories. Their conclusion drew on the following "think/no-think paradigm":

- In the first, preliminary phase, subjects are trained to learn some cue-target word pairs.
- In the experimental phase, which could be repeated several times, they were shown the cue. For some words they had to think about the target ("think" condition) while for others they had to suppress thoughts about the target ("no think" condition).
- In the final phase the subject's memory of the target associated with each cue was assessed.

Anderson and Green (2001) have shown that the "think/no think" instruction had a significant effect on recall. The "think" instruction made retrieval of the target easier, while the "no think" instruction made it harder. This effect was moreover strengthened by the number of repetitions of the experimental phase. Encouraging retrieval with rewards did not affect those results. Anderson and Green (2001, p. 303) concluded that their study "provides a mechanistic basis for the voluntary form of repression (suppression) proposed by Freud". Schacter (2001) noticed that this was at least overstated, given that the targets learned and suppressed were not emotionally loaded. The "think/no think" has however recently been applied to targets containing negative emotional content, showing that "when cognitive control mechanisms are directed toward suppression, their effect on memory representations is heightened for emotional compared with neutral stimuli" (Depue et al., 2006, 2007).

This is not, however, enough to vindicate a motivational unconscious. These studies indeed show that a subject can voluntarily forget something but they do not show that the suppressed memory remains in an unconscious form. It could have simply been destroyed. As Kihlstrom (2002) has pointed out "there was no evidence presented of persisting unconscious influence of the suppressed items. And there was no evidence that the 'amnesia' could be 'reversed'." Such an evidence would be however required in order to show the existence of a motivational unconscious.

This failure has to do with the fact that memory suppression is a form of apparent self-deception which can straightforwardly be explained in deflationary terms as a form of *diachronic* self-deception, and which does not require, for that reason, the presence of unconscious mental states.

3.2. MOTIVATED CONFABULATION

The term "confabulation" was coined in the beginning of the century to refer to the condition of some amnesic patients who seemed to invent, in "good faith," narratives about themselves and about the world, and showed little concern about the plausibility

of their stories. In his monograph devoted to the topic, Hirstein (2006) offers the following description of an encounter with a confabulating patient:

> A neurologist enters a hospital room and approaches an older man sitting up in bed. The neurologist greets him, examines his chart, and after a brief chat in which the man reports feeling fine, asks him what he did over the weekend. The man offers in response a long, coherent description of his going to a professional conference in New York City and planning a project with a large research team, all of which the doctor writes down. The only problem with this narration is that the man has been in the hospital the entire weekend, in fact for the past three months. What is curious is that the man is of sound mind, yet genuinely believes what he is saying. When the doctor informs him that he is mistaken, he replies, "I will have to check with my wife about that," then seems to lose interest in the conversation.

A similar phenomenon has since been found to affect some patients suffering from anosognosia for plegia, Anton syndrome, and hemineglect. In an experimental setting, it could also affect the patients whose *corpus callosum* has been severed ("split-brain" patients). In all those cases, the situation is somehow similar. Information coming from a given source is abnormally scarce (short-term memory in Korsakoff's syndrome, proprioception in anosognosia for hemiplegia, vision in Anton syndrome, sensory information exchanged through the corpus callosum in split-brain patients, etc.). When information from this source is solicited, the patients seem to fill the gaps by concocting stories.

Some have argued finally, that normal subjects are also prone to confabulation when asked to justify an attitude the cause of which is not normally available to introspection (Nisbett and Wilson, 1977; see also Bortolotti and Cox, 2009, p. 957–958, Dennett, 1992). This has led some to widen the meaning of confabulation. We will say that while confabulation in the narrow sense can only affect patients suffering from amnesia, confabulation in the wide sense is associated with other pathologies as well as with normal subjects. In this section, we will only deal with confabulation in the narrow sense.

If explanations of confabulation commonly appeal to cognitive factors, recent studies have suggested that they should invoke motivational factors and even M-unconscious states as well.

3.2.1. Cognitive explanations

Many have noticed that classical cognitive explanations of confabulations suffer from important shortcomings. It is customary to divide cognitive explanations of mental disorders into three categories:

- bottom-up, "empiricist" theories explain it in terms of an experiential deficit,
- top-down, "rationalist[12]" deficits explain it in terms of an executive function ("rationality") deficit,
- "two-factor[13]" accounts explain it in terms of both kinds of deficits.

As confabulation (in the narrow sense) is connected to a memory deficit, purely rationalist accounts are implausible. We are left with two groups of explanations. An inability to retrieve the relevant memories or to properly situate them in time, associated with a normal gap-filling mechanism could plausibly give rise to something like confabulation. However, it is well known that amnesia does not always give rise to confabulation. This has led some to postulate, in addition to the memory deficit, an executive dysfunction leaving patients unable to properly assess the hypotheses they form to fill the gaps in their memories. The occurrence of such a deficit is plausible since frontal sites responsible for the cognitive control and monitoring of thoughts are frequently damaged in confabulating patients (Fotopoulou et al., 2004; Turnbull et al., 2004). This two-factor explanation might not be entirely satisfactory, however, since it cannot account for confabulation in the wide sense, which does not seem to require an executive dysfunction. It might be answered that we should not expect that a single causal mechanisms underlies all forms of confabulations. Another problem with this account is that it resembles the so-called two-factor explanations of delusions. It might not, accordingly, be in a position to differentiate between delusions and confabulations (some actually welcome this consequence; see Bortolotti and Cox, 2009, pp. 955–956).

3.2.2. Motivational influences

The objections against cognitive explanations of confabulation, even though they might not be totally conclusive, might invite one to look for alternative explanations. There are reasons to think that motivational factors can also play a role in confabulation. It has long since been observed that the content of confabulations is often self-serving. This clinical observation has started to gain empirical confirmation. Using independent raters Fotopoulou et al. (2008b) showed that confabulating patients distorted their previous experiences in ways that are more pleasant than healthy and amnesic controls (Solms, 2000; Fotopoulou et al., 2004 and Turnbull et al., 2004 provide convergent data).

Among others, Marks Solms took the motivational account of confabulation to vindicate some Freudian ideas (Solms, 2000; Turnbull and Solms, 2007). It does not however vindicate the Freudian unconscious. For all those studies show is that the false beliefs of the confabulating patients fit their desires. They constitute a case of *apparent self-deception*. They give us no reason, however, to think that this apparent self-deception should be explained in terms of M-unconscious mental states rather than in a deflationary way, as a form *motivationally biased* belief formation (a form of mere wishful thinking). They do not show that the patients have a conscious true belief that contradicts their conscious false belief nor an unconscious motive to acquire this conscious false belief. In order to do that, an account of confabulation would have either to give direct empirical evidence of conflicting unconscious mental states or, at least, to show that

[12] I borrow this terminology from Campbell (2001) who uses it to classify the different accounts of delusions.

[13] The term is from Davies et al. (2001) who coined it to describe their own account of monothematic delusions.

because the patients' apparent self-deception involve selectivity and inner conflict, it cannot be accounted for in a deflationary way.

3.2.3. Lack of selectivity

As far as confabulation in the narrow sense is concerned, there is no direct empirical evidence of such unconscious states available yet. Furthermore the patients' apparent self-deception does not exhibit the features that would rule out an explanation in terms of a motivational bias. First, the content of confabulation is not *specific enough* to suppose that it is intentional: their confabulations tend to bear on their past but this can be explained by their amnesia, which facilitates their wishful thinking on this topic. This is indeed how Fotopoulou et al. (2008a) explain the phenomenon: "both confabulation and its motivated content result from a deficit in the control and regulation of memory retrieval, which allows motivational factors to acquire a greater role than usual in determining which memories are selected for retrieval. To this extent, the self-enhancing content of confabulation could be explained as a neurogenic exaggeration of normal self-serving memory distortion."

3.2.4. Lack of inner conflict

Second, the patients do not exhibit the signs of an inner conflict. Their behavior has not been shown to be ambivalent. Hirstein (2006, p. 180) remarks that they have an hyporesponsive or unresponsive autonomic system and suggests that they would not show the emotional reaction which is characteristic of lying. Conjoined with the lack of selectivity this apparent lack of inner conflict tends to show that anosognosia is a form of facilitated wishful thinking rather than a form of self-deception.

3.3. ANOSOGNOSIA FOR PLEGIA

Patients with anosognosia for plegia deny their paralysis. They can also produce narratives "covering" their ailment, and thus confabulate in the wide sense. They constitute an interesting case because (i) the explanation of their pathology seems to require, like that of confabulation in the narrow sense, the appeal to motivational factors, (ii) but they *do* sometimes show signs of an inner conflict. I tackle this second point first.

3.3.1. Clinical evidence of inner conflicts

It has been observed clinically that many patients seem to be somehow cognizant of their condition. Ramachandran (2009) for example describes the following case (notice that he appeals to a divided self approach to resolve the apparent inconsistency of the patient):

> An intelligent and lucid patient I saw recently claimed that her own left arm was not paralyzed and that the lifeless left arm on her lap belonged to her father who was "hiding under the table." Yet when I asked her to touch her nose with her left hand she used her intact right hand to grab and raise the paralyzed hand – using the latter as a "tool" to touch her nose! Clearly somebody in there knew that her left arm was paralyzed and that the arm on her lap was her own, but "she" – the person I was talking to – didn't know.

Other patients seem to avoid bimanual tasks. Ramachandran also described what he interpreted as "reaction formations" betraying the patients' knowledge. One of his patients for example claimed "I can't wait to go back to two-fisted beer drinking" (Ramachandran and Blakeslee, 1998, p. 139). Another wrongly claimed "I tied the shoelace *with both my hands*" (Ramachandran and Blakeslee, 1998, p. 139).

3.3.2. Empirical evidence of inner conflicts

Several measures have been recently put forward to assess those apparent inner conflicts. Marcel et al. (2004) asked patients to estimate their bimanual abilities in the first-person form ("In your present state how well, compared with your normal ability, can you. . .") and in the third-person form ("If I were in your present state, how well would I be able to. . ."). They concluded that 15–50% of the patients suffering from anosognosia with a right-brain damage showed greater overestimation of their abilities when asked in the first-person rather than in the third-person. Marcel et al. (2004) also obtained indirect acknowledgments of the deficit by asking the patients about their weak limb "in a series of different ways (e.g., in a tentative, confidential manner): "Is it ever naughty? Does it ever not do what you want?" While most patients suffering from anosognosia expressed incomprehension or replied negatively, five right-brain damaged patients replied affirmatively. One patient responded conspiratorially "Oh yes! In fact, if it doesn't do what I want, I'm going to hit it." As they recognize, "It is not yet clear (. . .) what the relevant aspect of [the] question was: type of description, confidentiality of interaction, emotional tone, or infantile speech."

Bisiach and Geminiani (1991) report that some patients may deny their plegia while spontaneously avoiding bimanual actions whereas others explicitly grant their plegia but spontaneously engage in bimanual actions. Such dissociations between implicit and explicit knowledge of plegia were recently tested by Cocchini et al. (2010) on a group of 30 right-brain damaged patients. They found that two patients exhibited a selective implicit anosognosia and eight patients exhibited a selective explicit anosognosia. Finally, Nardone et al. (2007) tested two patients using an attentional-capture paradigm. Unlike non-anosognosic controls, they showed a higher simple reaction time when a word related to movement was presented with the target, suggesting again, an implicit awareness of their deficit.

There is thus consistent evidence that some patients suffering from anosognosia, unlike patients suffering from confabulation in the narrow sense, have conflicting mental states relative to their plegia. We shall also see that, like the latter, their condition seems to result from motivational influences.

3.3.3. The explanatory argument for motivational influences

The denial of illness is a false belief which is self-serving and probably pleasant. But this is not enough to show that it is influenced by motivational factors for it could be self-serving so to speak, by chance (I believe that the weather is sunny today, which is actually conform to my desire, but there was no causal influence of my desire on my belief). One reason to affirm that it is influenced by motivational factors appeals to the insufficiency

of non-motivational explanations of anosognosia. Like the classical explanations of confabulation (in the narrow sense) classical, cognitive, explanations of anosognosia divide roughly into two groups. According to the empiricist accounts, the patient would deny his plegia because of a proprioceptive loss, of an hemispatial neglect, or because of a more subtle somatosensory deficit, such as an impairment of the feedforward mechanism responsible for the monitoring and control of limb movements (see Vuilleumier, 2004 for a review). Those empiricist theories are unsatisfactory because double dissociations between those somatosensory deficits and anosognosia have been observed, and because, more deeply, those somatosensory deficits cannot explain why the patient maintains his denial despite the contrary evidence that other reliable sources (vision, testimony, etc.) keep providing[14]. In response some theorists have proposed a two-factor theory which posits an executive problem, such as confusion or inflexibility. Those executive problems would however have to be very important and many data suggest that they are actually absent in some patients (Marcel et al., 2004; Vuilleumier, 2004).

The insufficiency of both kinds of cognitive theories might be seen as vindicating the causal role of motivational factors in anosognosia (Levy, 2008). Even when they are conjoined with experiential and executive deficits however, motivational factors will not be sufficient to explain all cases of anosognosia. Prigatano et al. (2010) reports for example a patient suffering from anosognosia and blindness. Suffering from hemineglect, she was no more aware of his visual experience than of her proprioceptive experience in her left hemifield. She nevertheless denied her plegia but acknowledged her blindness. It would be *ad hoc* to suppose that she had more motivations to deny her plegia than her blindness. All this suggests that the shortcomings of traditional explanations are often due to the fact that anosognosia is not a unitary phenomenon with a single underlying cause, rather than to the neglect of motivational factors.

3.3.4. The direct argument for motivational influences

This is not to say that motivational factors do not play a role in the explanation of some forms of anosognosia. Indeed some of the data collected by Marcel et al. (2004, p. 33) precisely suggest that they do. They observed, for example that "men over-estimated their abilities on car driving more than women and only in the first-person version of the question." This overestimation might be motivated by self-esteem, car driving abilities being more relevant to the latter for men than for women. The following observation is also interesting in that respect. After giving the patients a bimanual task, Marcel et al. (2004) asked them to explain their failure. It appeared that a significant portion of patients gave bizarre answers (e.g., "I should use a robot," or "My arm has a cold") suggesting both an implicit knowledge of their handicap and "a desperate attempt at defense against [explicit] acknowledgment" Marcel et al. (2004, p. 34).

Let us take stock. The false beliefs of some patients with anosognosia seem to arise from motivational factors. Those patients are apparently self-deceived. Moreover, they have conflicting attitudes toward their plegia, which suggests that this case of apparent self-deception is better explained in terms of a present intention to deceive oneself rather than in a deflationary way, as a form of diachronic self-deception or as a mere motivationally biased belief. This gives us an empirical argument for the existence of a motivational unconscious. The force of the latter should not be over-estimated though. It is actually a variant of the general argument:

1. Exclusion of deflationary approaches. Cases of apparent self-deception which involve selectivity and inner conflict are better explained in a non-deflationary way.
2. Empirical observation. Anosognosia is a form of apparent self-deception which involves an inner conflict.
3. General explanatory principle. A case of apparent self-deception which is better explained in a non-deflationary way involves a form of motivational unconscious.

The only difference is that the existence of the relevant case of apparent self-deception is supported by empirical evidence concerning anosognosia (3) whereas it is supported by folk-psychological observations about normal behavior in the original argument. The progress is meager given that the general principle is far more plausible in the case of normal behavior than it is in the case of anosognosia. The general explanatory principle, it should be reminded, rests on the assumption that subjects are rational enough to detect obvious inconsistencies between their conscious attitudes. It is not obvious, however, those patients who suffer from anosognosia and are relevantly self-deceived – those who exhibit a form of inner conflict – satisfy this rationality assumption.

More fundamentally, the general explanatory principle rests on various premises that might be deemed controversial and we should expect an empirical confirmation to give us independent support for its conclusion. It rested, for example, on the claim that apparently self-deceived patients seem to be deceiving themselves and that we should by default save this appearance. Such premises, again would not weigh much against some empirical evidence which favors alternative explanations of the apparent self-deception afflicting the patients. Some influential researchers have indeed argued for such alternative explanations. Hirstein (2006) has argued for an explanation in subpersonal terms. Marcel et al. (2004) have argued for an explanations in personal but *conscious* terms. I will argue that if the empirical evidence gathered so far seems to exclude Hirstein's account, it also seems to favor Marcel's neo-dissociationist account over those which picture the repressed states as unambiguously unconscious.

3.3.5. Subpersonal self-deception?

Hirstein (2006) has claimed that cases of apparent self-deception in patients suffering from anosognosia could be explained without having to postulate that the subject knows about his plegia and intends to forget about it. Hirstein's point rests mainly on the fact that the selectivity of the patient's confabulations can be explained

[14] Hohwy and Rosenberg (2005) and Pacherie (2008b) appeal to a form of vicious epistemic circle to explain the fixity of other delusional beliefs. The idea is that the subject could, without irrationality, use his delusional belief in order to assess the conflicting evidence. It is not obvious to me that such a strategy can generalize to the case of anosognosia for plegia.

without positing an intention to deceive oneself (Hirstein, 2006, p. 230) and on the patients' attenuated autonomic response, which would betray an absence of inner conflict. Although he does not address the issue, Hirstein could countenance some of the evidence we have reviewed so far, which indicates a form of inner conflict, by claiming that even though some parts of the patient's brain represent his plegia and control his belief that he is healthy, the subject himself, as opposed to his brain, does not believe, unconsciously or otherwise, that he suffers from a paralysis and does not intend, consciously or otherwise to believe that he is healthy. He could claim, in other words, that the patient's "deceptive intention" and "knowledge" are subpersonal.

3.3.6. Conscious self-deception?

I take it, however, that the results of Marcel et al. (2004) support the claim that some patients do represent their condition at a personal level. They show that the patients can, when they are asked in the proper way, *report* their handicap (cf. Levy, 2008, pp. 235–238). Those striking results, however, tend to show too much. They tend to show not only that the patients have genuine personal level attitudes toward their plegia, but also that they have *conscious* attitudes toward their plegia. They tend to show, more precisely that the patients consciously believe that they are paralyzed. After all, reportability is a good enough evidence for consciousness! Marcel and his colleagues accordingly opt for an interpretation of apparent self-deception among patients suffering from anosognosia in terms of *conscious* self-deception:

> Some patients seemed to show a genuine dissociation within awareness according to the manner or viewpoint of the question. Such a split in consciousness, often referred to as a dissociative state, has been said to occur in a variety of circumstances. (Marcel et al., 2004, p. 33)

Such an option might seem problematic because other measures reveal that the patients consciously believe that they are *not* paralyzed. Marcel et al. (2004, p. 33) suggests that this "is only a problem if we assume a strict singularity of consciousness and a logical consistency within any one segment of consciousness. That is, genuine dissociations and conflicts within awareness may be possible. (Marcel et al., 2004, p. 33)." I will come back to the question of the "strict singularity of consciousness" when we will put forward a neo-dissociationist approach to self-deception, according to which repressed state are in some important sense conscious even though they are not well accessible to the subject for rational control (see Sections 6.1–6.2). For now, it is enough to realize that the patients could simply fail to be rational enough to monitor their inconsistencies and that they could accordingly believe *consciously* both that they do and that they do not suffer from a paralysis.

4. CONCEPTUAL OBJECTIONS TO A M-UNCONSCIOUS

It might be no accident that two leading groups of researchers who recognize the conflicting nature of some forms of anosognosia nevertheless reject an explanation of this conflict in terms of unconscious personal level states. Doing so, they implicitly get round two conceptual objections that make the very notion of a motivational unconscious problematic. I will actually argue that if those two objections do not make M-unconscious states impossible, they make it at least quite different from what we have been looking for through most empirical studies. They would thus also contribute to explain why the empirical studies we have scrutinized have not provided some decisive evidence for a motivational unconscious.

Those two conceptual objections are related with the second and the third constraint on a motivational unconscious respectively. They rely on the tension between those and the first constraint. The first objection claims that *personal level* mental states like motives must necessarily be conscious (see Section 4.1). The second objection claims that one could not repress a mental state unless he is conscious of that state, so that repression is *a priori* impossible (see Section 4.2). In that section, I will review those two objections and some answers to them that I deem inadequate. I will then put forward my own neo-dissociationist answer.

4.1. A FIRST CONCEPTUAL OBJECTION: CAN PERSONAL LEVEL STATES BE UNCONSCIOUS?

Typical personal level mental states are conscious (at least as far as occurrent mental states are concerned). Typical unconscious brain states are not personal level mental states. Some researchers have suggested that unconscious mental states of a personal level would simply be chimerical. Wittgenstein for example considered that personal level accounts could only explain our actions by providing *motives* or *reasons* for them (he often seems to assimilate motives and reasons) and that unconscious mental states could not provide such motives or reasons for our actions but only mere causes for them[15]. Jaspers (1997) also considered that unconscious, as opposed to merely "unnoticed," mental states could not figure in personal level explanations[16].

4.1.1. Challenging the conscious-unconscious distinction

Even if it were acknowledged that some personal level mental states could be unconscious, there would still be a problem of criterion. On what ground should we judge that an unconscious state is a personal level mental state? Consciousness and reportability are the most common and the most natural criteria for the personal level, but those are useless when it comes to assess the existence of *unconscious* personal level mental states. Levy (2008) argues at length that the empirical evidence supports the claim that patients suffering from anosognosia really do "acknowledge," *at a personal level*, their plegia. To counter Hirstein (2006)'s contention that this "acknowledgment" is only subpersonal because it is not available to them, he has to argue that availability comes in degree and that this acknowledgment is somehow available to them. The idea

[15]"The difference between a reason and a cause is brought out as follows: the investigation of a reason entails as an essential part one's agreement with it, whereas the investigation of a cause is carried out experimentally. (. . .) Of course the person who agrees to the reason was not conscious at the time of its being his reason. But it is a way of speaking to say the reason was subconscious. (. . .) It is a confusion to say that a reason is a cause seen from the inside. A cause is not seen from within or from without. It is found by experiment (Wittgenstein, 1982, p. 11)."

[16]Unlike Wittgenstein who seemed to believe that reasons or motives could not be causes Jaspers believed that personal level explanations were also causal explanations. Both agreed however that unconscious mental states could only provide the mere causes of an action, and never "make sense" of it.

seems to be that at a certain level of availability we can say that the patient has *some consciousness* of his condition and that he accordingly acknowledges it at a personal level even though he is not *fully conscious* of it. I think that this is actually a good strategy, and we shall see that a defender of the M-unconscious might have to adopt a similar neo-dissociationist strategy, arguing that M-unconscious state are not totally unconscious. But unless more is said, it might seem like an *ad hoc* maneuver to counter the claim of those who contend that either a state is sufficiently available to the subject to be a personal level state, and it is conscious, or it is not but it is unconscious. We will come back to that point later (see Section 5.4.2).

4.1.2. Challenging the personal-subpersonal distinction

Alternatively, instead of making the border between the conscious and the unconscious murkier, one might try to challenge the distinction between personal and subpersonal mental states. Bermúdez (2000) has for example argued that personal level explanations are not autonomous in that some subpersonal level mental states can, and should figure, in some explanations that "make sense" of an action. He mentions, among others, explanations of the behavior of blindsights (Weiskrantz, 1986), or explanations of skillful behavior (some subpersonal motor processes can explain why the tennis champion managed to put the ball just here and not a few inches further, behind the line; see also Pacherie, 2008a). I agree with the general claim that some subpersonal states can figure in explanations that make sense of an action. I am not sure however, that such subpersonal states can, in those explanations, play the role of M-unconscious states. In particular, if the notion of M-unconscious is consistent, there should be some M-unconscious *intentions* capable alone of making sense of some actions, namely, the intentions to deceive oneself. It is dubious, however, that subpersonal intentions can alone, make sense of actions, or even that there can be such a thing as a subpersonal intention (an intention of whom?). This brings us to the second conceptual problem.

4.2. THE FUNDAMENTAL PROBLEM OF THE M-UNCONSCIOUS: IS IT POSSIBLE TO REPRESS AN UNCONSCIOUS STATE?

We saw that the notion of M-unconscious allow us to solve the paradoxes of self-deception. It is however exposed to what one might call a "revenge problem[17]". The problem, to put it briefly, is that *in order to act on something intentionally I must be conscious of the thing I act on*. Otherwise I could not have a sufficient *control* of that thing for my action to count as intentional. This is true even if we understand, as we do, "intentional" in a minimal sense which does not involve anything like deliberations. When John shoots his wife in anger, if his action is intentional, he must be conscious of his wife and of the fact that she is in a position to be shot. In the same way, if I close the door of a conference I am attending to and that in doing so I prevent a very noisy man from entering the room, we will not say that I *intentionally* prevented this man from entering unless I was aware of this man. More generally, if I am not conscious of X and X happens because of me, then it happens so to speak "blindly" rather than intentionally: as a consequence of some automatic, subpersonal process in me, or as a mere collateral effect of some other thing that I do intentionally. This *principle of conscious agency* connecting intentional action and consciousness applies to repression. It yields that the subject must be conscious of the states he represses, which contradicts the claim that they are unconscious. This revenge problem[18] is, I take it, the fundamental problem of the motivational unconscious. If M is a typical M-unconscious state, an M-unconscious state, that is, which results from repression,

1. The repression of M is intentional (The M-unconscious is a Motivated Unconscious)
2. In order to repress something *intentionally* one must be conscious of that thing (Principle of conscious agency)
3. The subject is not conscious of M[19] (The M-unconscious is unconscious)

(1) and (3) are definitional. (2), the principle of conscious agency, is widely endorsed by psychologists who consider intentional control as a measure of consciousness (Debner and Jacoby, 1994). It has been defended on intuitive grounds: intentional actions, as opposed to automatic behaviors and mere collateral effects must not be "blind". Here is a more cautious argument connecting consciousness and action through the notion of control:

- If a subject acts intentionally on X, X must be under his *intentional control*.
- If X is under the *intentional control* of a subject, that subject must be conscious of X.

[17] Sartre (1976, p. II) famously argued that the unconscious posited by Freud was thwarted by some kind of a revenge problem. However, the revenge problem he delineated depends on the problematic assumption that Freud adopted a divided mind approach to apparent self-deception, and it is accordingly misses its target. Drawing on Sartre and on Freud himself, Fingarette (1983) addresses an objection to Freud's classical theory of the unconscious which is quite close to the one I frame in that subsection.

[18] This revenge problem actually stems from a what might be considered a third paradox of self-deception. This third paradox, which I have called the agentive paradox elsewhere (Billon, unpublished) goes as follows:

1. In order to repress someone's belief intentionally I must be able to control this belief.
2. In order to control the belief of someone I must be aware of the fact that he has this belief.
3. So in order to repress one of my beliefs intentionally I must be aware of the fact that I have the belief.

This paradox, like the others will not be solved by the claim that the subject is divided between various centers of agency, for such a division makes the motivation for repressing the belief mysterious (see Section 2.3.2). If the M-unconscious provides a solution to the first two paradoxes of self-deception, the latter seems thwarted by this agentive paradox.

[19] There is actually a small loophole for one might be conscious of a mental state without being conscious of being in that state. This seems to happen for example, in patients suffering from thought insertion (Billon, forthcoming). I take it, however, that to repress a sate intentionally, one must be not only conscious of that state but also of being in that state. Moreover, standard cases of repression (unlike maybe so-called "projections") do not involve the misattribution of a mental state.

5. AN AFFECTIVE, NEO-DISSOCIATIONIST PICTURE OF THE M-UNCONSCIOUS

If the concept of M-unconscious is to be consistent, the argument (1–3) which has it that M-unconscious states are both conscious and unconscious must rest on an equivocation. This hypothesis is indeed quite plausible. Consciousness is a famously ambiguous and slippery notion (Chalmers, 1996; Siewert, 1998; Block, 2004). As Block (2004) puts it, "there are a number of very different "consciousnesses". (. . .) These concepts are often partly or totally conflated, with bad results." The fundamental problem of the M-unconscious might be one of those "bad results." It could indeed be escaped if "conscious" did not have the same meaning in (2) and in (3). Considering that consciousness and agency are related by the principle of conscious control, distinctions in forms of consciousness are likely to mirror distinctions in forms of intentional control of agency. I will indeed argue that we should distinguish phenomenal consciousness and access-consciousness on the one hand (Section 5.1) and impulsive and rational forms of intentional control on the other hand (Section 5.2). Repression is, I will argue, an impulsive process, and repressed states are conscious in the phenomenal sense even if the subject is not access-conscious of them and cannot accordingly control them rationally. This will allow us to solve the fundamental problem of the M-unconscious. We shall see that this will allow us to solve the first conceptual problem as well.

5.1. THE AMBIGUITY OF CONSCIOUSNESS

5.1.1. Access-consciousness (A-consciousness)

One of the difficulties surrounding the task of defining consciousness is that the word applies to different kinds of thing. We can talk of

- conscious creatures (this man is conscious),
- of conscious mental states (his perception of the grass is conscious),
- and of conscious contents (he is conscious *of* the grass).

As the example makes clear those senses are related. In particular a subject is conscious of R if he has a conscious state with the content R. The risk of confusion is even bigger when the conscious content is itself a mental state, like when I am conscious of my desire to go to the sea, for in that case the desire which is a conscious content can also be a conscious state. In order to avoid ambiguities, I will always use the locution "conscious *of*" to refer to conscious contents.

Consider now a perceptual state which represents that the grass is green in front of me. We can say that this perceptual state is conscious (or that I am conscious of the grass being green) to mean that I can report that the grass is green, that I can reason on this fact and infer that it rains often in this part of the country, etc. This sense of consciousness underlies reportability. It is called A-consciousness.

Access-consciousness (A-consciousness) If a subject S has a state X which represents R, we say that X is A-conscious, or that S is A-conscious of R, if R is broadcast for free use in the *rational control* of behavior (I adapt this definition from Block, 2004).

A-consciousness marks a degree of availability, or a quality of access. We can indeed classify degrees of availability depending on the kind of behavior the available information can yield:

a. Minimally available information. An information is minimally available, when it is only represented at a subpersonal level and does not normally influence the subject's overt behavior. Brain potentials reveal that patients suffering from prosopagnosia covertly recognize faces (Renault et al., 1989), this recognition however seems to be, at least in certain patients, only minimally accessible (Bruyer, 1991).
b. Subliminal information 1. At a higher degree of accessibility, the information influences the subject's behavior in subtle ways. One can think for example of the priming effects of masked stimuli (Marcel, 1983; Dehaene and Naccache, 2001; Kouider et al., 2006).
c. Subliminal information 2. At a slightly higher degree, the subject can report the information in forced choice situations (think of the visual information available to blindsights; Weiskrantz, 1986).
d. Subliminal information 2'. The information might also be available in such a way that although the subject cannot report it spontaneously, he could have done so if his attention had not been impeded. This is the case of stimuli which are shown during an attentional blink (Sergent and Dehaene, 2005; Dehaene et al., 2006 calls such subliminal stimuli *preconscious*).
e. Non-categorical spontaneous report. The information might be available in such a way that the subject can spontaneously report it but only in a non-categorical way. For example, a subject who feels depressed might be unable to categorize his feeling as depression and yet be able to discriminate when this feeling changes or disappears.
f. Categorical spontaneous report. At a higher level, the subject is also able to categorize the information.
g. Memory. To store it in memory.
h. Predicting and planning. And to use it for making predictions, for planning, etc. The typical supraliminal perception of "medium sized dry goods" for example allows to do all those things.

The degree of availability to which A-consciousness corresponds is the degree of availability which allows *rational control*. In Blocks's official definition "The "rational" is [just] meant to rule out the kind of automatic control that obtains in blindsight (Block, 2004)," so rational control and A-consciousness would probably start at (e). For reasons that will appear later[20], I think it is better to understand "rational" in a more substantial sense, and to consider that A-consciousness and the kind of rational control that characterizes it start at (f). It will be useful to give a name to the kind of

[20]In a nutshell, the principle of conscious agency reveals a conceptual connection between consciousness and intentional control. We will see that the category of intentional control divides in two subcategories: impulsive and rational control. The distinction between two forms of consciousness seems to mirror that distinction between two forms of intentional control, but in order to acknowledge this, we must adopt this more substantial reading of "rational" in the definition of A-consciousness.

rational control and A-consciousness involved in normal supraliminal perceptions (h): we will say that the, subjects are *robustly A-conscious* of their contents and that he can use the information they provide for *highly rational control*.

5.1.2. Phenomenal consciousness (P-consciousness)

Whatever the precise degree of availability it requires, A-consciousness will conceptually differ from phenomenal consciousness (P-consciousness). My phenomenally conscious states are those states there is something it is like to be in (Nagel, 1974), those states that have a subjective or a qualitative character. They are what we call subjective experiences. While access-consciousness is tied to rationality, phenomenality is tied *affectivity*[21]. It should be noted however, that even if affects and emotions are paradigmatic of phenomenal states, perceptions, episodes of inner speech are also phenomenally conscious. Some have even argued that abstract thoughts could be conscious in this sense as well (Siewert, 1998; Pitt, 2004). Just like for A-consciousness we can attribute P-consciousness both to states and to their contents. Thus, if a subject has a subjective experience X of R we can say that he is P-conscious of R or that X is P-conscious.

Phenomenal consciousness (P-consciousness) We say that a subject's mental state is P-conscious if it is a subjective experience. We say that subject S is P-conscious of R if he has a subjective experience of R.

It has often been noted that because of their subjectivity, P-conscious states seemed to involve a form of implicit or marginal reflexivity (Kriegel, 2004). What it's like to have an experience is always what it's like *for me* to have an experience because, to put it simply, I experience my subjective experiences. In what follows it will be useful to remember this suggestion to the effect that we are indeed always P-conscious *of* our P-conscious states, be it in am implicit and marginal manner.

5.1.3. The (un-)consciousness of M-unconscious states

We can now come back to the fundamental problem of the M-unconscious. In what sense should "unconscious" be taken in (3). In what sense should a subject be unconscious of his M-unconscious states. If a subject has an M-unconscious state M, he should be unable to report M. More broadly, he should be unable to use the information that he is in M for taking rational decisions. The subject should accordingly be A-*un*conscious of M. We can thus rephrase (3) as:

3'. S is A-unconscious of M (**M-unconscious is unconscious**)

If we are to defuse the second conceptual objection we must show that there is a sense of consciousness, say consciousness*, such that

i. consciousness* is a form of consciousness in an intuitive sense.
ii. if S is conscious* of M, M is under his *intentional control*.
iii. S can be conscious* of M without being A-conscious of M. S can be conscious* of M, that is, without M being under his *rational control*.

(iii) seems to stand in tension with (ii) and (i). I will argue that this tension is only apparent and that we have independent reasons to believe that P-consciousness can satisfy (i–iii). I will indeed show that there is an important distinction between two forms of intentional control that mirrors the distinction between A-consciousness and P-consciousness:

- There is a form of control, impulsive control, that is intentional without being rational.
- This intentional form of control is associated, as the principle of conscious agency requires, with a form of consciousness, namely P-consciousness.
- A subject can be P-conscious of something he is not A-conscious of.

In the picture that will emerge, M-unconscious states are typically affective states that the subject experiences viscerally even though he is not A-conscious of them. He can control them impulsively, but not rationally and repression is thus an impulsive process.

5.2. THE AMBIGUITY OF INTENTIONAL CONTROL

Why should we posit sub-rational forms of intentional control? According to a tradition that can be traced back to Hume and that has been, at least until recently, quite dominant, all intentional actions suppose a form of rational control. Doing something intentionally is doing it *for a reason*, and the intention of the action is precisely the reason for carrying it. In other words, all our motives or intentions would be equally rational. Such a rationality of intentional actions presupposes that we always chose the action which satisfies our dearest desires, the one, that "maximizes our preferences." This *Humean conception of motivation* and action is shared by classical rational choice theory, by classical cognitive psychology and by the very influential writings of Donald Davidson on action.

5.2.1. The problem of impulsivity

The Humean conception is confronted, however, with what one might call the problem of impulsivity. Consider a drug addict who sincerely desires to stop taking drugs, but nevertheless goes and gets one more dose. He seems to be acting intentionally against his best reasons. He might even explicitly realize that what he is doing goes against his dearest desires and his interest. If this is indeed a correct description of his behavior, then some impulsive actions are intentional actions which are not carried for a reason. They are intentional actions which do not involve rational control[22].

5.2.2. Dissolving the problem of impulsivity

Confronted with the problem of impulsive actions researchers have adopted three different stances. Some have tried to dissolve

[21] Here again, even though I literally accept Block's official definition, I interpret it in a slightly different manner by insisting on the connection between P-consciousness and affectivity. This is quite commonsensical I suppose, as witnessed by the common idea that what phenomenally unconscious creatures, like robots, would paradigmatically lack is emotions (Spielberg, 2001).

[22] As already noted it is traditional to distinguish impulsivity from akrasia in the strict sense. Akratic actions, it is said, can supposedly be done cold-bloodedly. I agree that some akratic actions seem cold-blooded. I am not sure however that they really are so and I tend to be skeptical about this distinction.

the problem by claiming that appearances notwithstanding, "impulsive actions" are either rational or not actions at all. With respect to addictive behaviors both variants of this "dissolutionist answer" have been defended. The economist (and Nobel prize winner) Gary Becker has for example claimed that addictive behavior could, under certain hypotheses, be explained within the framework of rational choice theory (Yaffe, 2001). At the other end of the spectrum, some neuroscientists have claimed that addiction is "nothing but" a brain disease and that addictive behaviors has to be understood in subpersonal terms, as automatic happenings, rather than in personal terms, as genuine actions of the subject (Leshner, 1997). Both variants are unsatisfying. Against Becker, drug-addicts often experience an inner conflict, which the theory of "rational addiction" does not, and probably cannot, account for. Against the theory of addiction as a brain disease, addictive behaviors do indeed exhibit the features of intentional actions. First, the subject does not behave automatically, he has *some control* over his behavior, in only of a limited form. As Mele (2004) emphasizes, there are always some motives that could prevent someone from carrying a given impulsive action. If a drug-addict is seriously threatened, say with a gun, or if he has to brave his biggest phobia to get a dose, he will normally refrain from going and getting it. Second, and relatedly, even if the subject cannot react rationally to it, he is *conscious* of the craving that moves him. Cravings are indeed paradigmatically things one experiences, things that are P-conscious (they might also be A-conscious, but we shall see that this is not necessary).

5.2.3. Solving the problem of impulsivity by renouncing the Humean theory of motivation

The problem of impulsivity should then invite us to reject the Humean theory of motivation and opt for what we might call a *Platonic conception of motivation*[23] According to such a theory, motivation is not monolithic. It has actually two distinct dimensions, which correspond to two forms of control and which are connected, as required by the principle of conscious agency, to two forms of consciousness (**Table 2**):

- Rational actions, or actions made for a reason, are controlled by *rational evaluation* which rely on *A-conscious* reasons.
- Impulsive actions, are controlled by visceral or *affective appraisal* which rely on *P-conscious affects* or visceral factors.

Such a theory revives the classical distinction between reason and passions, volitions and inclinations or desires and drives. It has been defended against the dominant Humean conception by a few philosophers (Watson, 1975; Schapiro, 2009). It has overall been promoted by studies in neuroeconomy. In his seminal work, Loewenstein (1996) has urged to dissociate rational from visceral factors in the explanation of actions. The later are defined by their direct hedonic impact, and their direct influence on our behavior. They are very little sensitive to reasons and can have disproportionate effects. These analyses have recently been confirmed by the discovery of two neuroanatomically distinct systems responsible for rational and visceral decisions respectively (Sanfey et al., 2003, 2006).

Table 2 | The Platonic conception of motivation.

Intentional actions		Automatic behaviors
Actions for a reason	**Impulsive actions**	
Personal level	Personal level	Subpersonal level
E.g., singing, sitting-down	E.g., Phobic reaction, addictive behavior	E.g., digestion, sneezing
Require A-consciousness	Require P-conscience	Do not require consciousness

Some intentional actions are not made for a reason, they are accomplished impulsively. Being intentional actions, they require a form of consciousness, but unlike rational actions which require A-consciousness, they only require P-consciousness.

5.3. P-CONSCIOUSNESS WITHOUT A-CONSCIOUSNESS

We could then solve the second conceptual problem by claiming that repression is an intentional action of the subject which is impulsive and non-rational. He can control a state M he represses because he experiences it, because he is P-conscious of it. As P-conscious states are subjective, which implies, as we saw, an implicit reflexivity, he might even be P-conscious of M just in virtue of the fact that M is itself P-conscious.

For that solution to work, however, we would have to claim that although he experiences it, he is not A-conscious of it. Some might object that one can always report his experiences and that one cannot accordingly be P-conscious of something without being A-conscious of it. In a series of publications, Ned Block has defended that "P-consciousness overflows A-consciousness" (Block, 2001, 2005, 2007). His arguments rest on sophisticated conceptual and methodological considerations (Block, 2007). A simpler, and quite commonsensical, argument is available to us[24].

5.3.1. Diffuse feelings

When I experience something, I feel it, for sure, but what I feel might not be enough to allow the rational control of my behavior. Consider a case in which you are diffusely anxious because (say) you will be undertaking a big voyage soon. You experience this diffuse anxiety and if it suddenly went away, you would notice a change. In that sense, you can report this anxiety, connected with your imminent travel. But you might be unable report it under that mode of presentation, categorically that is. You might be unable to tell that you feel *anxious* (categorical recognition), let alone

[23] For lack of space, I will neglect what might be called the "minimal solution" to the problem of impulsivity. The minimal solution tries to solve problem by modifying the rational choice theory but without renouncing the Humean theory of motivation as such. Such a solution is associated with the works of George Ainslie, it explains impulsive action by a form of hyperbolic discount of the future (Ainslie, 1975, see Yaffe, 2001; for a short introduction and critique of the theory).

[24] This commonsensical argument might be unavailable to Block because, as we noted (see Section 5.1.1), he understands A-consciousness in a slightly more inclusive way.

that this anxiety is caused by your apprehension of an imminent voyage (etiological recognition). You might accordingly be totally unable to use what you experience for the rational control of your actions. This contrasts drastically from cases of regular supraliminal P-conscious perceptions of medium sized dry goods. When you P-consciously perceive R, you can normally recognize that your conscious state is a perception of R (categorical recognition) and that it is caused by R (etiological recognition). This is why our perceptions can normally feed the rational, and indeed, the highly rational, guidance of our actions.

Such subjective experiences of which the subject is not A-conscious might be quite widespread. In an important article, Lambie and Marcel (2002) devote many pages to such situations which they characterize as situations of emotion experience without second-order awareness (this terminology is consistent with mine: the A-consciousness *of* an emotion is a form of *second-order* awareness, for emotions are mental states). They describe a few case reports involving people apparently unable to report their fear experience but they also note that the empirical data on "repressors" (Weinberger, 1990) and on alexithymia (Lane et al., 1997) might be interpreted as evidence for their claim. More to the point, on Clark (1986)'s influential account, panic attacks would result from an inability of the subject to be A-conscious of the anxiety he experiences. Unable to categorize it as such, the subject would mistake its bodily effects for the symptoms of an illness or of a heart attack. Lambie and Marcel (2002) also remark that patients with "anger disorders" (Kassinove and Sukhodolsky, 1995), who show high or chronic levels of anger "often feel in some sense that "the world is against them", and (. . .) many of them lack awareness of their anger (Eckhardt and Deffenbacher, 1995), and are having treatment only because employers or loved ones have encouraged them to do so (Lambie and Marcel, 2002)." Drawing on the cognitive treatments of anger disorders, they claim that the patients experience a form of anger which they are not A-conscious of. They conclude, strikingly, that "there is every reason to suppose that the same occurs for a range of other emotions as for anger. The only reason that unrecognized anger is highlighted is that its effects cause other people to make the individual see a therapist. Unrecognized happiness or love does not have the same effect (Lambie and Marcel, 2002)."

5.3.2. Cold–hot empathy gaps as affect blindness

These difficulties to become A-conscious of some P-conscious states that we experience seem to be linked to the fact that those P-conscious states are hot, visceral, affective, P-conscious states – they are some of those states that are paradigmatically P-conscious and that motivate impulsive actions. There is indeed ample evidence, quite generally, that their hot character makes those affective states indistinct and little visible to A-consciousness. Even when we *can* categorically recognize and report some of those states, it appears that we cannot remember them well nor use them well for making predictions and planning. In other words, even when we are A-conscious of them, our A-consciousness of them is not of a robust form. The availability of the information they provide tends not to be good enough to serve highly rational reactions. It has been shown that we systematically underestimate the impact of visceral factors in our decisions and in our explanations and our predictions of behavior. Loewenstein (1996, pp. 277–278) for example reports that when asked to make a non-binding choice, expectant women tend to state a desire to eschew anesthesia but reverse their decision when they start to experience pain. "The reversal of preference, he notes, was observed not only for women giving birth for the first time, but also those who had previously experienced the pain of childbirth; experience does not seem to go very far in terms of enhancing one's appreciation for future pain." Such "cold–hot empathy gaps" have since then been found in a variety of situations, in intrapersonal and interpersonal settings, regarding past, present as well as imagined affects (Wilson and Gilbert, 2003; Loewenstein, 2005; Van Boven and Loewenstein, 2005).

We could summarize those findings on empathy gaps by saying that just as our P-conscious affective appraisals are little sensitive to A-conscious reasons, conversely, our A-conscious rational evaluations are not highly sensitive to our P-conscious affects. This finding is important because it tends to make of A-consciousness and rational evaluations on the one hand, and of P-consciousness and affective appraisals on the other hand two partly independent psychological systems: a cold system and a hot systems that have different ways of learning (Phelps, 2005), different dynamics, and different functions. The partial independence of those systems explains many of our irrational behaviors, and in particular, our tendency to repeat the same mistakes without ever seeming to learn from experience (Gilbert, 2006). They could account, very broadly, for what Haybron (2008) has called the (widespread) "pursuit of unhappiness". This finding, finally, echoes many developments in the so-called dual-process psychology Evans (2008), not to mention Freud's distinction between primary and secondary processes.

5.4. THE AFFECTIVE M-UNCONSCIOUS

Let us take stock. We saw that the notion of M-unconscious is faced with two conceptual problems. We saw that the second of these problems, which we dubbed the fundamental problem of the M-unconscious could indeed be solved thanks to the distinction between two forms of intentional agency and two correlated forms of consciousness.

5.4.1. The affective M-unconscious and the fundamental problem of the M-unconscious

With these distinctions in mind we can indeed rephrase the fundamental problem of the M-unconscious as an argument to the effect that M-unconscious states must be P-conscious states the subject is not A-conscious of, which result from an impulsive repression, and which are probably hot affective states (**Figure 1**):

1. The repression of M is intentional (The M-unconscious is a motivated unconscious)
2. If a subject represses a state intentionally, he must be conscious of that state (Principle of conscious agency)
   - If a subject represses a state intentionally he represses it *either for a reason or impulsively* (The ambiguity of intentional control)
   - If one represses a state *for a reason*, he must be A-conscious of that state (Rational control and A-consciousness)

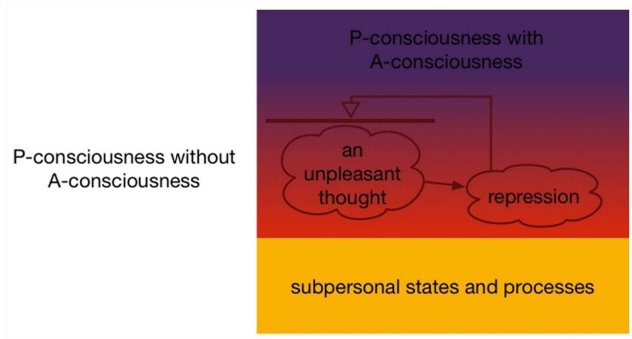

FIGURE 1 | The affective understanding of the M-unconscious. M-unconscious thoughts are affective P-conscious thoughts. Repression, which prevents the subject from being A-conscious of those thoughts, is itself a P-conscious, impulsive process of which the subject is not A-conscious.

- If one represses a state *impulsively*, he must be P-conscious of that state (Impulsive control and P-consciousness)

3. The subject is not A-conscious of M (The M-unconscious is unconscious)

3′. The subject must be P-conscious of M but not A-conscious of M, and the repression must be impulsive [from (2) and (3)]

4. If the subject is P-conscious of M but not A-conscious of M and the repression is impulsive, M is probably a hot affective state [from (3') and the data on states of which the subject is P-conscious but not A-conscious].

As an example we can return to the case of the Mary who deceived herself into believing that she is not suffering from cancer. According to this affective understanding of the M-unconscious, she would experience some intense fear of dying subsequent to her diagnosis. This fear would push her into rejecting the diagnosis, into avoiding his physician, and into considering her fear of dying as something else, maybe as an indeterminate anxiety, or even as a mere bodily phenomenon.

It should be noticed that this notion of an affective M-unconscious which is experienced but of which the subject is not A-conscious is not altogether absent from Freud's writings. It is according to Freud one sense of "unconscious": the sense of unconscious as applied to affects. In a text in which he struggles to define a precise and unified concept of the unconscious, Freud notes that even though "it is of the essence of an emotion that we should be aware of it, i.e., that it should become known to consciousness (. . .) in psycho-analytic practice we are accustomed to speak of unconscious love, hate, anger, etc., and find it impossible to avoid even the strange conjunction "unconscious consciousness of guilt", or a paradoxical "unconscious anxiety" (Freud, 1974, p. 3000)". Unconscious affects, he says, are affects that are separated from their representation and which the subject can accordingly experience even though he is not A-conscious of their representational content. Being thus separated from their representation, they do not convey information about their category and their cause, and they cannot be used for rational control (Freud, 1974, pp. 3000–3003). I have just argued that this sense of the unconscious, which is tied to phenomenality and affectivity, is the only consistent sense[25].

5.4.2. *The affective M-unconscious and the first conceptual problem of the M-unconscious*

What about the first conceptual problem? We evoked some solutions that were not entirely satisfying. Those solutions relied on the claim that despite their unconscious character, M-unconscious states could be considered as personal level states because they had some degree of availability. The problem with those solutions was that they sounded like *ad hoc* maneuvers. The notion of P-conscious states of which the subject is not A-conscious allows us to argue quite independently that there are some states which, although they are not sufficiently available to the subject for him to be A-conscious of them, are nevertheless personal level states. They are personal level states because they are conscious in the phenomenal sense and because conscious states, in any intuitive sense, are always personal level states.

6. CONCLUSION

Reviewing some recent work relevant to self-deception and the M-unconscious we have provided a general theoretical argument for its existence. We have assessed some influential empirical studies often mentioned as a vindication of the M-unconscious. We have finally noted that the notion of M-unconscious suffered from two important conceptual problems, which plague the empirical investigation of the M-unconscious, and which could only be solved at the price of an important shift in understanding. Where does this shift in its understanding leave us, relative to the general argument and to the empirical evidence of M-unconscious states?

6.1. REASSESSING THE GENERAL ARGUMENT: M-UNCONSCIOUS AND APPARENT SELF-DECEPTION

The M-unconscious was a good candidate to solve the paradoxes of apparent self-deception and to provide an explanation of this vexing phenomenon. The conceptual problems which afflict it could however be considered like revenge paradoxes showing that the notion is not that well suited for the job. The solution we proposed to these revenge paradoxes, in terms of affective M-unconscious, actually allows the M-unconscious to play the role it was designed for. But it has us picture these states in a more Janetian and in a less, or less straightforward, Freudian way. They would be states that are *dissociated* from those states the subject is A-conscious of, but states that are P-conscious nevertheless rather that totally unconscious.

The only viable M-unconscious approach to apparent self-deception would thus be what could be called a *neo-dissociationist approach*. It should be noted that a few recent works from Anthony Marcel and his colleagues have revived such an approach to psychic, and in particular emotional, phenomena (cf., e.g., Lambie and Marcel, 2002).

[25]The same claim is made by the phenomenologist Henry (1985), in his important **Genealogy of Psychoanalysis**.

REASSESSING THE EMPIRICAL EVIDENCE

As for the empirical evidence for the M-unconscious, our shift in understanding seems to impose a disappointing conclusion. The most promising evidence was obtained in studies of patients from anosognosia. These studies, we concluded tended to favor a neo-dissociationist picture of (some forms of) this condition according to which the patients are not unambiguously unconscious of their plegia. It is however far from clear that they favor the neo-dissociationist picture of the M-unconscious defended here, for it is not clear that patients suffering from anosognosia really experience their plegia without being A-conscious of it. Lambie and Marcel (2002) suggest that something close to the opposite happens with a subset of those patients (they would be A-conscious of their plegia in a detached way, but they would not experience it directly, and they would not feel accordingly feel concerned about it). As already mentioned, Hirstein (2006) emphasizes that they have a reduced autonomic response (I do not know, however, of any empirical study testing this).

In any case, we have seen that patients suffering from anosognosia are very varied, and it is not excluded that a subgroup of patients might indeed know, so to speak affectively and M-unconsciously, about their plegia (maybe some of the patients that Cocchini et al., 2010 classify as having a selective explicit anosognosia).

Our shift in understanding also opens novel perspectives in the empirical study of the M-unconscious. If M-unconscious states are P-conscious, they might be easier to spot than it is usually assumed.

Also some results interpreted as refuting the M-unconscious might need to be reconsidered. For example a study by Giesbrecht et al. (2007) showed that high dissociative tendencies were associated with higher skin conductance (SC) response to an emotional video. They concluded that

> Our SC findings are difficult to reconcile with the view that dissociative experiences constitute a defense mechanism that allows an individual to psychologically withdraw from the impact of the situation. Such a mechanism should reduce emotional engagement, thereby leading to lowered SC responses.

If the M-unconscious is an affective M-unconscious and M-unconscious are affective states the subject experience, however, we should not expect defensive function to reduce emotional engagement, but only, at best, to reduce the emotional engagement associated with the A-consciousness of one's emotions, what might be called the second-order emotional engagement. More generally, if this account of the M-unconscious is on the right track, it is a mistake to think about flattened affect and about the disorders in which it occurs, like depersonalization disorders (Sierra, 2009), as the possible result of repression.

Finally, and most importantly, this novel understanding provokes us to go and look further in the direction of affective sciences, dual-process psychology and neuroeconomy in order to ascertain the existence of a M-unconscious. We did find, in those fields, evidence for P-conscious states of which the subject is not A-conscious. This is not enough to show that the lack of A-consciousness is motivated, but it might be a good start.

ACKNOWLEDGMENTS

I would like to add a thanks to Marie-Noelle Jeu, Kassy Hayden as well as Michel Crubelier and Linda Brakel for friendly comments on an earlier draft.

REFERENCES

Ainslie, G. (1975). Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychol. Bull.* 82, 463.

Anderson, M., and Green, C. (2001). Suppressing unwanted memories by executive control. *Nature* 410, 366–369.

Bargh, J., Chen, M., and Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *J. Pers. Soc. Psychol.* 71, 230.

Bermúdez, J. (2000). Personal and subpersonal: a difference without a distinction. *Philos. Explor.* 3, 63–82.

Bermùdez, J.-L. (2000). Self-deception, intentions, and contradictory beliefs. *Analysis* 60, 309–319.

Bisiach, E., and Geminiani, G. (1991). "Anosognosia related to hemiplegia and hemianopia," in *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues*, eds G. Prigatano and D. Schacter (Oxford: Oxford University Press), 17–39.

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition* 79, 197–219.

Block, N. (2004). "Some concepts of consciousness," in *Philosophy of Mind, Classical and Contemporary Readings*, ed. D. Chalmers (New York, NY: Oxford University Press. Abriged version of Block (1995)), 206–218.

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 9, 46–52.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.*

Bonanno, G., Keltner, D., Holen, A., and Horowitz, M. (1995). When avoiding unpleasant emotions might not be such a bad thing: verbal-autonomic response dissociation and midlife conjugal bereavement. *J. Pers. Soc. Psychol.* 69, 975.

Bortolotti, L., and Cox, R. E. (2009). 'faultless' ignorance: strengths and limiitations of epistemic definitions of confabulation. *Conscious. Cogn.* 18, 952–965.

Bruyer, R. (1991). Covert face recognition in prosopagnosia: a review. *Brain Cogn.* 15, 223–235.

Campbell, J. (2001). Rationality, meaning and the analysis of delusion. *Philos. Psychiatr. Psychol.* 8, 89–100.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York, NY: Oxford University Press.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *J. Philos.* 78, 67–90.

Clark, D. (1986). A cognitive approach to panic. *Behav. Res. Ther.* 24, 461–470.

Cleeramns, A., Destrebcqz, A., and Boyer, M. (1998). Implicit learning: news from the front. *Trends Cogn. Sci. (Regul. Ed.)* 2, 406–416.

Cocchini, G., Beschin, N., Fotopoulou, A., and Sala, S. D. (2010). Explicit and implicit anosognosia or upper limb motor impairment. *Neuropsychologia* 48, 1489–1494.

Davidson, D. (1983). "Paradoxes of irrationality," in *Philosophical Essays on Freud*, Chap. 17, eds R. Wollheim and J. Hopkins (Cambridge: Cambridge University Press), 289–305.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davidson, D. (2004). *Problems of Rationality*, Vol. 4. New York, NY: Oxford University Press.

Davies, M., Breen, N., Coltheart, M., and Langdon, R. (2001). Monothematic delusions: towards a two-factor account. *Philos. Psychiatr. Psychol.* 8, 133–158.

Debner, J. A., and Jacoby, L. L. (1994). Unconscious perception: attention, awareness, and control. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 304–317.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci. (Regul. Ed.)* 95, 529–534.

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

Dennett, D. (1969). *Content and Consciousness*. London: Routledge.

Dennett, D. (1992). "The self as a center of narrative gravity," in *Self and Consciousness: Multiple Perspectives*, eds F. Kessel, P. Cole, and D.

Johnson (Hillsdale: Erlbaum, États-Unis), 275–288.
Depue, B., Banich, M., and Curran, T. (2006). Suppression of emotional and nonemotional content in memory. *Psychol. Sci.* 17, 441.
Depue, B., Curran, T., and Banich, M. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science* 317, 215.
Eckhardt, C., and Deffenbacher, J. (1995). *Diagnosis of Anger Disorders.*
Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278.
Fazio, R., and Williams, C. (1986). Attitude accessibility as a moderator of the attitude–perception and attitude–behavior relations: an investigation of the 1984 presidential election. *J. Pers. Soc. Psychol.* 51, 505.
Fingarette, H. (1969). *Self-Deception.* Berkeley: University of California Press.
Fingarette, H. (1983). "Self-deception and the 'splitting of the ego,'" in *Philosophical Essays on Freud,* Chap. 13, eds R. Wollheim and J. Hopkins (Cambridge: Cambridge University Press), 212–227.
Fotopoulou, A., Conway, M. A., Solms, M., Tyrer, S., and Kopelman, M. (2008a). Self-serving confabulation in prose recall. *Neuropsychologia* 46, 1429–1441.
Fotopoulou, A., Conway, M. A., Tyrer, S., Birchall, D., Griffith, P., and Solms, M. (2008b). Is the content of confabulation positive? An experimental study. *Cortex* 44, 764–772.
Fotopoulou, A., Solms, M., and Turnbull, O. (2004). Wishful reality distortions in confabulation: a case report. *Neuropsychologia* 42, 727–744.
Freud, S. (1974). *The Complete Psychological Works of Sigmund Freud,* Standard Edn. Ed. and trans. James Strachey (London: Hogarth Press), 24.
Gardner, S. (2006). "The unconscious," in *The Cambridge Companion to Freud*, Chap. 5, ed. J. Neu (Cambridge: Cambridge University Press), 136–160.
Giesbrecht, T., Geraerts, E., and Merckelbach, H. (2007). Dissociation, memory commission errors, and heightened autonomic reactivity. *Psychiatry Res.* 150, 277–285.
Gilbert, D. (2006). *Stumbling on Happiness.* New York, NY: Knopf.
Hassin, R., Uleman, J., and Bargh, J. (2005). *The New Unconscious.* New York, NY: Oxford University Press.
Haybron, D. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being.* New York, NY: Oxford University Press.
Henry, M. (1985). *Généalogie de la Psychanalyse.* Épiméthée, PUF, Paris.
Hirstein, W. (2006). *Brain Fiction – Self-Deception and the Riddle of Confabulation.* Cambridge: MIT Press.
Hohwy, J., and Rosenberg, R. (2005). Unusual experiences, reality testing, and delusions of alien control. *Mind Lang.* 20, 141–162.
Jaspers, K. (1997). *General Psychopathology,* 2 Edn. (Trans. J. Hoenig and Marian W. Hamilton). Baltimore: JHU Press.
Kassinove, H., and Sukhodolsky, D. (1995). *Anger Disorders: Basic Science and Practice Issues.* Washington, DC: Taylor & Francis.
Kihlstrom, J. (1987). The cognitive unconscious. *Science* 237, 1445.
Kihlstrom, J. (2002). No need for repression. *Trends Cogn. Sci. (Regul. Ed.)* 6, 512.
Kouider, S., Dehaene, S., Jobert, A., and Bihan, A. L. (2006). Cerebral bases of subliminal and supraliminal priming during reading. *Cereb. Cortex* 17, 2019–2029.
Kriegel, U. (2004). Consciousness and self-consciousness. *Monist.* 87, 185–209.
Lambie, J., and Marcel, A. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychol. Rev.* 109, 219.
Lane, R., Ahern, G., Schwartz, G., and Kaszniak, A. (1997). Is alexithymia the emotional equivalent of blindsight? *Biol. Psychiatry* 42, 834–844.
Laplanche, J., and Pontalis, J.-B. (1968). *Vocabulaire de la psychanalyse,* 2 Edn. Paris: Presses Universitaires de France.
Leshner, A. (1997). Addiction is a brain disease, and it matters. *Science* 278, 45.
Levy, N. (2008). "Self deception without thought experiments," in *Delusions and Self Deception: Affective and Motivational Influences on Belief-Formation*, eds T. Bayne and J. Fernández (New York, NY: Psychology Press), 227–242.
Loewenstein, G. (1996). Out of control: visceral influences on behavior. *Organ. Behav. Hum. Decis. Process* 65, 272–292.
Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health Psychol.* 24, S49–S56.
Marcel, A. (1983). Conscious and unconscious perception: an approach to the relation between phenomenal experience and perceptual processes. *Cogn. Psychol.* 15, 197–237.
Marcel, A. J., Tegnêr, R., and Nimmo-Smith, I. (2004). Anosognosia for plegia: specificity, extension, partiality and disunity of bodily unawareness. *Cortex* 40, 19–40.
Mele, A. (1997). Understanding and explaining real self-deception. *Behav. Brain Sci.* 20, 127–134.
Mele, A. (2004). "Volitional disorders and addiction," in *The Philosophy of Psychiatry: A Companion,* Chap. 5, ed. J. Radden (New York, NY: Oxford University Press), 78–88.
Mele, A. R. (2001). *Self-Deception Unmasked.* Princeton, NJ: Princeton University Press.
Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450.
Nardone, I., Ward, R., Fotopoulou, A., and Turnbull, O. (2007). Attention and emotion in anosognosia: evidence of implicit awareness and repression? *Neurocase* 13, 438.
Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84, 231–259.
Nussbaum, M. (1994). *The Therapy of Desire.* Princeton, NJ: Princeton University Press.
Pacherie, E. (2008a). The phenomenology of action: a conceptual framework. *Cognition* 107, 179–217.
Pacherie, E. (2008b). "Perception, emotions and delusions: revisiting the capgras' delusion," in *Delusions and Self-Deception: Affective Influences on Belief Formation,* eds T. Bayne and J. Fernandez (New York, NY: Psychology Press), 107–126.
Pears, D. (1987). "The goals and strategies of self-deception, in *The Multiple Self, Studies in Rationality and Social Change,* Chap. 2, ed. J. Elster (Cambridge: Cambridge University Press), 59–78.
Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R., and Frith, C. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science* 316, 904.
Phelps, E. (2005). "The interaction of emotion and cognition: the relation between the human amygdala and cognitive awareness," in *The New Unconscious*, eds R. Hassin, J. Uleman, and J. Bargh (New York, NY: Oxford University Press), 61–76.
Pitt, D. (2004). The phenomenology of cognition, or what is it like to think that p? *Philos. Phenomenol. Res.* 69, 1–36.
Prigatano, G. P., Matthes, J., Hill, S. W., Wolf, T. R., and Heiserman, J. E. (2010). Anosognosia for hemiplegia with preserved awareness of complete cortical blindness following intracranial hemorrhage. *Cortex* doi: 10.1016/j.cortex.2010.11.013. [Epub ahead of print].
Ramachandran, V. S. (2009). Self-awareness: The last frontier. *Edge.org.* Available at: http://www.edge.org/3rd_culture/rama08/rama08_index. html
Ramachandran, V., and Blakeslee, S. (1998). *Phantoms in the Brain: Human Nature and the Architecture of the Mind.* New York, NY: William Morrow.
Renault, B., Signoret, J.-L., Debruille, B., Breton, F., and Bolgert, F. (1989). Brain potentials reveal covert facial recognition in prosopagnosia. *Neuropsychologia* 27, 905–912.
Rorty, A. O. (1988). "The deceptive self: liars, layers, and lairs," in *Perspectives on Self-Deception*, Chap. 1, eds B. P. McLaughlin and A. O. Rorty (Berkeley: University of California Press), 11–28.
Sanfey, A. G., Loewenstein, G., McClure, S. M., and Cohen, J. D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends Cogn. Sci. (Regul. Ed.)* 10, 108–116.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758.
Sartre, J. (1992). *Being and Nothingness: A Phenomenological Essay on Ontology.* New York, NY: Washington Square Press.
Sartre, J.-P. (1976). *L'Être et le Néant. Folio.* Paris: Gallimard.
Schacter, D. L. (2001). Suppression of unwanted memories: repression revisited? *Lancet* 357, 1724–1728.
Schapiro, T. (2009). The nature of inclination. *Ethics* 119, 229–256.
Schmidt, L., Palminteri, S., Lafargue, G., and Pessiglione, M. (2010). Splitting motivation. *Psychol. Sci.* 21, 977.
Sergent, C., and Dehaene, S. (2005). Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. *J. Physiol. Paris* 98, 374–384.
Sierra, M. (2009). *Depersonalization: A New Look at a Neglected Syndrome.* Cambridge: Cambridge University Press.
Siewert, C. (1998). *The Significance of Consciousness.* Princeton: Princeton University Press.

Solms, M. (2000). A *psychoanalytic* perspective on confabulation. *Neuropsychoanalysis* 2, 133–138.

Spielberg, S. (2001). *Artificial Intelligence*. Movie based on the short story of Brian Aldiss "Super-toys last all summer long".

Turnbull, O., Jenkins, S., and Rowley, M. (2004). The pleasantness of false beliefs: an emotion-based account of confabulation. *Neuropsychoanalysis* 6, 5–16.

Turnbull, O. H., and Solms, M. (2007). Awareness, desire, and false beliefs: Freud in the light of modern neuropsychology. *Cortex* 43, 1083–1090.

Uleman, J. (2005). "Introduction: becoming aware of the new unconscious," in *The New Unconscious*, eds R. Hassin, J. Uleman, and J. Bargh (New York, NY: Oxford University Press), 3–15.

Van Boven, L., and Loewenstein, G. (2005). "Empathy gaps in emotional perspective taking," in *Other Minds: How Humans Bridge the Divide Between Self and Others* (New York, NY: Guilford Press), 284–297.

Vuilleumier, P. (2004). Anosognosia: the neurology of beliefs and uncertainties. *Cortex* 40, 9–17.

Watson, G. (1975). Free agency. *J. Philos.* 72, 205–220.

Weinberger, D. (1990). "The construct validity of the repressive coping style," in *Repression and Dissociation: Implications for Personality Theory, Psychopathology, and Health*, ed. Jerome L. Singer (Chicago, IL: University Of Chicago Press), 337–386.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. New York, NY: Oxford University Press.

Wilson, T., and Gilbert, D. (2003). Affective forecasting. *Adv. Exp. Social Psychol.* 35, 345–411.

Wittgenstein, L. (1982). "Conversations on freud; excerpt from 1932-3 lectures," in *Philosophical essays on Freud*, eds R. Wollheim and J. Hopkins (Cambridge: Cambridge University Press), 1–11.

Yaffe, G. (2001). Recent work on addiction and responsible agency. *Philos. Public Aff.* 30, 178–221.

Yalom, I. (1980). *Existential Psychotherapy*. New York, NY: Basic Books.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 August 2011; paper pending published: 15 August 2011; accepted: 23 August 2011; published online: 27 September 2011.

Citation: Billon A (2011) Have we vindicated the motivational unconscious yet? A conceptual review. Front. Psychology ***2****:224. doi: 10.3389/fpsyg.2011.00224*

This article was submitted to Frontiers in Psychoanalysis and Neuropsychoanalysis, a specialty of Frontiers in Psychology.

Copyright © 2011 Billon. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.