

行政院國家科學委員會專題研究計畫 成果報告

自我與現象意識(第3年) 研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC 97-2410-H-004-154-MY3
執行期間：99年08月01日至100年09月30日
執行單位：國立政治大學英國語文學系

計畫主持人：藍亭

計畫參與人員：碩士班研究生-兼任助理人員：薛旭任
碩士班研究生-兼任助理人員：張哲虹
碩士班研究生-兼任助理人員：張凱琪
碩士班研究生-兼任助理人員：蘇偉誠
碩士班研究生-兼任助理人員：周映妤
大專生-兼任助理人員：黃品嘉
大專生-兼任助理人員：呂懷哲
博士班研究生-兼任助理人員：周怡岑
博士班研究生-兼任助理人員：沈映伶

報告附件：出席國際會議研究心得報告及發表論文

公開資訊：本計畫可公開查詢

中華民國 100 年 12 月 30 日

中文摘要：在心理學、神經科學尚未發展的六〇年代，哲學家要研究心理或心靈，僅能透過內省或純思考方式。美國哲學家 Sydney Shoemaker 在 1968 年提出「IEM (immunity to error through misidentification)」學說，他認為當「我」做為主體，只要透過內省來感知痛覺、觸覺等感官經驗，那必然是「我」的感受，不可能辨識錯誤。例如當一個人說自己牙痛的時候，我們並不會質疑對方「那真的是你的痛覺嗎？」此外，Shoemaker 主張此關係不僅維持我們的體感經驗，同時也作用於行動意識與視覺感知。

然而，本研究透過實際病例與實驗結果，證明此關係並非必然。以研究病患症狀為例，某些病患會將自己的肢體視為「他人的」，而產生主體與意識經驗的分離。舉例而言，某病患將自己的左手視為外甥女，實驗設計反覆觸碰其左手，當病患被告知其左手要被觸碰時，她表示並無感覺，只有當實驗者告知她「外甥女的左手要被觸碰了」，病患才有觸覺反應。此病例說明雖然「我」是主體，卻必須將意識經驗表徵為「他人的」，才能透過內省去體驗並恢復感知。

本研究以此成功推翻了長年以來廣為接受的 Shoemaker 學說 (IEM)，認為其學說僅能視為「假說」，並非任何情況下都能成立。本研究旨在以心理學、神經科學等研究方法來檢驗哲學問題，並期望達成以下目標：(1) 透過告知臨床醫生哪些問題應被問及，協助醫生更了解病患情況；(2) 設計實驗使我們更了解「自我」、「意識經驗」與「身體」之間的關係。

中文關鍵詞：自我, 意識, 心智歸屬, 體化, 自然主義, 神經哲學

英文摘要：The main contribution of my research over the past three years has been to show that conscious experiences can sometimes only occur if they are represented as belonging to someone other than self. The standard assumption is that when we introspectively know that a state like pain exists, we necessarily know that it is our pain. Most philosophers take this to be a tautology, a conceptual truth, or a datum. The most well-known articulation of this relationship between conscious states and self is Sydney Shoemaker 's' immunity to error through misidentification relative to the first-person pronoun' ('IEM'). David Rosenthal has also developed a version of this principle, which he calls 'thin immunity.' In a series of essays I have

argued that immunity principles should be regarded as hypotheses. Moreover, by adducing evidence from various pathological states (Anton 's Syndrome, Somatoparaphrenia, and Thought Insertion) as well as evidence from certain experimentally-induced illusions (the Rubber Hand Illusion and Full-Body Illusions) I have shown that IEM or IEM-like hypotheses are confronted by genuine counter-examples. First-person awareness that a conscious state is instantiated does not entail awareness that it is instantiated in oneself.

英文關鍵詞： Self, Consciousness, Mental Ownership, Embodiment, Naturalism, Neurophilosophy

行政院國家科學委員會專題研究計畫

自我與現象意識（第 3 年）
研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC97-2410-H-004-154-MY3
執行期間：99年08月01日至100年09月30日
執行單位：國立政治大學英國語文學系

計畫主持人：藍亭
Timothy Lane

備註：本計畫可公開查詢

中華民國 100 年 12 月 29 日

Table of content

Chinese Abstract.....	2
English Abstract.....	2
Key Words.....	3
Project Content:	3
Preface.....	3
Research Purpose.....	4
Literature Review.....	4
Methodology.....	14
Results and Discussion, Part I: Self-Consciousness.....	21
Self-Consciousness and Immunity.....	21
Mental Ownership and Higher-Order Thought.....	43
Higher-Order Thought and Pathological Self.....	48
A soft self and a hard core.....	54
Results and Discussion, Part II: Sleep Mentation.....	66
The threshold of wakefulness, the experience of control, and theory development.....	66
What subjective experiences determine the perception of falling asleep during sleep onset period.1. Introduction.....	69
Results and Discussion, Part III: Belief and Ethics.....	82
The ethics of false belief.....	82
Results and Discussion, Part IV: Anti-Individualism and Vision Science.	111
Results and Discussion, Part V: Partial and Whole Body Illusions.....	112
The malleability of self and body experiences.....	112
Self-specificity and mineness.....	112
Mental ownership and the rubber hand illusion.....	115
Results and Discussion, Part VI: The ethics of suicide research.....	115
Media Impact on Individual Suicidality-A proposal for an ethical neuroimaging study.....	115
Acknowledgments.....	116
References.....	116
Self-evaluation of project outcome.....	131

1. Chinese Abstract

在心理學、神經科學尚未發展的六〇年代，哲學家要研究心理或心靈，僅能透過內省或純思考方式。美國哲學家 Sydney Shoemaker 在 1968 年提出「IEM (immunity to error through misidentification)」學說，他認為當「我」做為主體，只要透過內省來感知痛覺、觸覺等感官經驗，那必然是「我」的感受，不可能辨識錯誤。例如當一個人說自己牙痛的時候，我們並不會質疑對方「那真的是你的痛覺嗎？」此外，Shoemaker 主張此關係不僅維持我們的體感經驗，同時也作用於行動意識與視覺感知。

然而，本研究透過實際病例與實驗結果，證明此關係並非必然。以研究病患症狀為例，某些病患會將自己的肢體視為「他人的」，而產生主體與意識經驗的分離。舉例而言，某病患將自己的左手視為外甥女，實驗設計反覆觸碰其左手，當病患被告知其左手要被觸碰時，她表示並無感覺，只有當實驗者告知她「外甥女的左手要被觸碰了」，病患才有觸覺反應。此病例說明雖然「我」是主體，卻必須將意識經驗表徵為「他人的」，才能透過內省去體驗並恢復感知。

本研究以此成功推翻了長年以來廣為接受的 Shoemaker 學說 (IEM)，認為其學說僅能視為「假說」，並非任何情況下都能成立。本研究旨在以心理學、神經科學等研究方法來檢驗哲學問題，並期望達成以下目標：(1) 透過告知臨床醫生哪些問題應被問及，協助醫生更了解病患情況；(2) 設計實驗使我們更了解「自我」、「意識經驗」與「身體」之間的關係。

2. English Abstract

The main contribution of my research over the past three years has been to show that conscious experiences can sometimes only occur if they are represented as belonging to someone other than self. The standard assumption is that when we introspectively know that a state like pain exists, we necessarily know that it is our pain. Most philosophers take this to be a tautology, a conceptual truth, or a datum. The most well-know articulation of this relationship between conscious states and self is Sydney Shoemaker's "immunity to error through misidentification relative to the

first-person pronoun” (“IEM”). David Rosenthal has also developed a version of this principle, which he calls “thin immunity.” In a series of essays I have argued that immunity principles should be regarded as hypotheses. Moreover, by adducing evidence from various pathological states (Anton's Syndrome, Somatoparaphrenia, and Thought Insertion) as well as evidence from certain experimentally-induced illusions (the Rubber Hand Illusion and Full-Body Illusions) I have shown that IEM or IEM-like hypotheses are confronted by genuine counter-examples. First-person awareness that a conscious state is instantiated does not entail awareness that it is instantiated in oneself.

3. Key Words

Self, Consciousness, Mental Ownership, Embodiment, Naturalism, Neurophilosophy

4. Project Content

A. Preface

This project holds out the promise of clarifying what kind of thing or process we are (i.e. what a self is)—in other words, that which is most distinctive about human beings—of what or whom gets lost in dementia, and with what or whom we are communicating when we converse with those suffering from PVS. We live on an island inhabited by thousands of PVS patients. But we have no way to reach out to them. I hope we can provide a way, a way that can then be applied to PVS patients in other parts of the world. Conceptual breakthroughs and new technology are making this possible. As a moral community we would be remiss were we to ignore this chance to communicate with those among us who are most isolated.

B. Research Purpose

(1) Theory-related purpose: sometimes conceptual arguments can take hold with such force that they become impediments to intellectual inquiry. I believe that this has been the case with immunity principles. The intent has been to remove this impediment—the immunity principles—to intellectual inquiry, by promoting an inter-animation of philosophical theory and empirical inquiry.

(2) Practical purpose: One, among many reasons, for pursuing this line of research, is to increase our understanding of standard conditions for the normal sense of self and its embodiment. By enhancing our understanding of embodiment in normal conditions, we can better project how this sense would vary under atypical circumstances, such as what might be experienced during periods of acceleration or deceleration, during periods of weightlessness, or when functioning at high altitudes.

(3) By extension from point we can better train those who rely upon prostheses or who work with tools in atypical environments. For the former, we would be enhancing their efficiency and their safety; for the latter, their quality of life.

C. Literature Review

1. Some Background:

The main purpose of this three-year project is to develop an idea that I first formally proposed last year, Mental Ownership Theory. This idea, however, has been percolating for several years. Mental Ownership Theory (MOT) concerns how we distinguish between mental states that belong to self and mental states that belong to others. I can know that a mental state (e.g. fear) exists, by “seeing” it in the eyes of a person standing before me; I can also know that fear exists via introspection. Under normal circumstances, in the former case we reliably attribute fear to someone else and, in the latter case, attribute it to self. This very natural tendency has led many philosophers and scientists to think that the difference between attributing fear to self and fear to others is strictly determined

by the mode-of-access: one via perception of something observable in the external world; the other, via something that can be known, “directly”, via introspection. But I (Lane and Liang 2009, 2010, 2011) have already demonstrated that the *access-distinction*—to a first approximation, introspection versus perception of the external world—is neither necessary nor sufficient for determining the self-other, the *ownership-distinction*. Mode-of-access is one among multiple factors that contribute to determining the *belongingness* of mental states.

Naturally, my point is not to deny that mode-of-access is highly relevant to determining ownership. But mode-of-access is not *the* determining factor. Were access not highly reliable in this respect, mobile creatures capable of mental states could not exist. Were we chronically confused about whether, say, body sensations belong to me, to someone else, or to no one, then it would be difficult to imagine circumstances that would allow for the existence of our species. If pain, or fear, or the visual experience of a rapidly approaching projectile exists, it is critical to know whether these belong to self or to someone else. In the former case, we had best be prepared to act. In the latter case, a more relaxed approach can be adopted—after all, it is not me who is in harm’s way.

It turns out to be the case that mode-of-access and ownership are just contingently related. This claim is highly counter-intuitive for most, if not all, people. The standard intuition is well articulated by Wittgenstein’s (1958: 66-67) famous rhetorical question: “...there is no question of recognizing a person when I say I have toothache (sic). To ask ‘are you sure it is *you* who have pains?’ would be nonsensical.” This intuition was then developed by many, most famously Shoemaker (1968).

Shoemaker (1996: 10) proclaimed that the relationship between a subject and an experience are as intimate as a “branch and a branch bending”: the “bending” cannot exist independently of the “branch.” Concepts of the relevant sort are tautologically (Shoemaker 1968: 563-564) related to one another, such that when we make a judgment like “I feel pain” we are aware that “one does, oneself, feel pain...one is, tautologically, aware, not simply that the attribute *feel(s) pain* instantiated, but that it is instantiated in oneself.” It “*cannot* happen that I am mistaken in saying ‘I feel pain’ because, although I do know of someone that feels pain, I am mistaken in thinking that person to be myself” (Shoemaker 1968: 557).

Cast in more formal terms, Shoemaker says of what he terms immunity to error “through misidentification relative to the first-person pronouns” (IEM) that

to claim that a statement “a is Φ ” might be erroneous through misidentification relative to the term “a” is to allow for the following possibility: “the speaker knows some particular thing to be Φ , but makes the mistake of asserting ‘a is Φ ’ because, and only because, he mistakenly thinks that the thing he knows to be Φ is what ‘a’ refers to.” But if Shoemaker has accurately identified a tautology, then mistakes of this type are not possible. Assuming that the ground of my judgment is introspection, whenever I say “I feel pain” it cannot be the case that I am mistaken in thinking that the person in pain is me.

Despite what strikes many—perhaps all of us—as a tautological relationship between concepts, it seems we are easily misled by analogies like “branch bending” and “branch.” It is the case that “bending” cannot occur without a “branch,” or some other similarly pliable object. But the relationship between a self (or a subject) and a mental state is importantly disanalogous to Shoemaker’s example. Although mental states do require brains (or some suitable brain substitute) to exist, it does not follow that a self and a mental state are related in the same way as “branch” and “bending.” As familiar cases of thought-insertion reveal (Stephens and Graham 2003), selves can be seriously confused as regards their relationship to mental states that are available to them via introspection.

During the first stage of developing MOT I have not, however, devoted much attention to thought-insertion. One reason for postponing treatment of thought-insertion is that during stage one I wanted first to expose the inadequacy of IEM and Gallagher (2000), along with others (e.g. Coliva 2000a and 200b), have previously defended IEM against objections raised that are based upon the conscious experiences of schizophrenics. Gallagher’s (2000: 231) main point is that a patient in a florid schizophrenic state claims only that he is not the “author” of thoughts; hence, they are felt to have been “inserted”. I do address Gallagher’s concerns (Lane and Liang, Forthcoming), but I began making the case against IEM with reference to a pathology of a different sort—somatoparaphrenia.

2. Somatoparaphrenia and IEM:

My first attempt to address the issue of immunity was not directly motivated by Shoemaker’s IEM. Instead, (Lane and Liang 2009 and 2010) it was motivated by a different version of immunity, one promulgated by Rosenthal, what he terms “thin” immunity (2002, 2004: 168-176 and 2005: 341-353). That was an extension of a more general critique (Lane and Liang 2008) of what Rosenthal

regards as an empirical theory of consciousness, Higher-Order Thought (HOT). But some aspects of MOT were developed while addressing a component of HOT—thin immunity—and while responding to Rosenthal’s (2010) defense of thin immunity against my criticisms. Here, however, I will concentrate just on those aspects that address the more well-known version of immunity, Shoemaker’s IEM. In both cases though, both IEM and thin immunity, the point of departure is a complex phenomenon, somatoparaphrenia.

Somatoparaphrenia (Vallar and Ronchi 2009) is a syndrome that is characterized by the sense of profound estrangement from parts of one’s body. It is typically found in patients who have suffered extensive right-hemisphere lesions (usually vascular). Lesions in the temporo-parietal junction are a common neural correlate of somatoparaphrenia, but deep cortical regions (e.g. the posterior insula) and subcortical regions (e.g. the basal ganglia) are also sometimes implicated. Somatoparaphrenia is also closely associated with proprioceptive impairment and often (not always) co-morbid with hemispatial neglect. Patients feel that a contralesional limb, most frequently the hand, seems not to belong to them; indeed, it often seems to belong to someone in particular, not uncommonly an acquaintance. The sense of disownership can be so vivid that even after recovery patients continue to describe the estrangement in factive, not metaphoric, language (Halligan 1995).

Somatoparaphrenia is occasionally accompanied not only by hemispatial neglect, it is also accompanied by tactile extinction (the loss of conscious tactile perception) in the estranged body part. Moro et al. (2004) demonstrated (for two cases) that by merely changing the position of the hands—moving the left hand across the midline of the body, over to the right-hand side—tactile sensation could be recovered. Even though tactile sensation could be so readily recovered, the sense of limb disownership was unchanged.

As regards my challenge to IEM, the most relevant case has been described by Bottini et al. (2002). A woman (FB) reported that her left hand belonged to her niece and that she (FB) felt no tactile sensations there. In FB’s case the lesion was subcortical, involving the basal ganglia, white matter underlying the insula, as well as other areas. But, importantly, the primary somatosensory area—which is critical to processing tactile sensation—was preserved. As Bottini et al (2002: 251) record: “F.B.’s spared *ability to perceive* tactile stimuli, provided that these were

referred to someone else's body, was evidently based on the survival of some elementary somatosensory cortical functions."

In a series of controlled tests, FB, while blindfolded, was advised that the examiner would touch *her* left hand; next the examiner would in fact touch the dorsal surface of FB's hand. Whenever this was done, FB said that she could feel no tactile sensations. When advised that the examiner was about to touch her *niece's* hand, however, upon actually being touched, she reported feeling tactile sensation. It is here that we begin to see the relevance of FB's case to IEM.

It should be born in mind that FB was in all other aspects cognitively sound. Moreover, in order to ensure that these tests would be reliable, catch trials—wherein FB was led to expect touches that were not forthcoming—were used. These trials were evenly distributed across three verbal warnings—I am going to touch your right hand, your left hand, and your niece's hand—and were administered in four sessions, two on one day, two on the next. It is of paramount importance to note that in not even one of 36 catch trials, 9 each per session, did FB respond incorrectly. In other words, when advised that she (or her "niece") would be touched, if no contact was made, FB always reported "no," no contact had been made.

When reflecting upon IEM in light of Bottini et al.'s findings, I allowed that most of Shoemaker's views as regards self, mental states, and conscious experience are true. But even when assuming (for the sake of argument) that Shoemaker's principal theses are true, we are left with an explanatory puzzle: why is it that when FB is expecting to be touched (on the left hand), she feels nothing, whereas when she expects that her niece will be touched there, she is able to report tactile sensation? Why, despite the experimental controls that are in place (e.g. blindfold and catch trials), is she able to judge that "her niece" has been touched? Typically to say (a) "I am going to touch your arm," implies (b) "I am going to touch you." It would be nonsensical to say (a) without implying (b). Likewise, when the doctor says "I am going to touch your niece's hand," she implies that "I am going to touch your niece." The concern here is not about *where* the sensation will be felt, but about *who* will feel the sensation. Pace the prototypical situations that motivate the Wittgenstein-Shoemaker intuition—it is *not* absurd to inquire as to whom is the subject of experience.

If we divide the experiment into two parts: FB expecting that she will be touched is Part 1 and FB expecting that her niece will be touched is Part 2. FB's

case should be regarded as directly relevant to IEM because she has been primed by the doctor to introspect. I argue that in Part 2 FB is misrepresenting her tactile sensation as belonging to someone else. In Part 2, from FB's first-person perspective, when introspecting on that tactile sensation, FB is misrepresenting herself, such that she is not the owner of the sensation. In a word, FB commits an error through misidentification regarding just *who* is the subject of the sensation.

It is then empirically possible for a subject, while introspecting a mental state (and thereby knowing that someone is undergoing that state), to be in error with regard to whom is experiencing that particular mental state. Admittedly this is counterintuitive. The Wittgenstein-Shoemaker intuition that to inquire of the person who introspects and reports a toothache whether it is indeed that person who has the ache strikes all of us as absurd. But empirical inquiry has ways of upsetting the apple cart: it would by no means be absurd to ask of FB whether it is she who has the tactile sensations, even though it is she who produces the introspectively-based report.

Notice that there is an important contrast here that calls for an explanation. We have a fact and a foil, the contrast between the two parts of FB's case. In Part 1 when FB is primed to introspect on what she experiences, she reports nothing; in Part 2, when she is primed to introspect on what her niece experiences, she reports tactile sensation.

To ignore this difference would be to ignore a significant explanatory problem. Because FB Parts 1 and 2 have similar histories, it is possible to ask sensible contrastive questions, questions which enable us to elicit causal differences (Lipton 1993: 217-219). And this is a possibility that is not permitted by IEM. In this case the essential difference between the two is whether FB represents herself as subject of the mental state. This issue, concerning first-person representation of just *who* the subject is, I refer to as *mental ownership*.

One might worry that FB merely reports feeling the sensation, when in fact she does not feel anything. But on this series of tests had there been no actual sensations, the reports would not have been made. First, recall that in FB's case the lesion was subcortical; her somatosensory cortex was preserved. So it is not surprising that she retained the capacity for experiencing tactile sensations. Second, FB's performance on catch trials was perfect. Included among the catch trials were instances for which she was told that her niece's hand was about to be touched, when in fact it was not touched. In these trials she never once made a

false report—neither on the first nor the second day of the experiments. Third, other standard procedures were in place to monitor FB’s sustained attention and reliability of response: for example, FB’s hand was touched in a randomized, fixed sequence, which was repeated in four sessions, on two separate days. Therefore, because she was blindfolded and because of the other controls that were in place, she could only have given an accurate report had she actually experienced the sensation of being touched.

Moreover, imaging studies of self-referential processing show that there is no reason to suspect that problems pertaining to mental ownership typically impair tactile processing. Northoff et al.’s (2006) analysis of many and varied studies that engage the “feeling of mineness” indicates that these experiences are subsumed by a set of commonly activated regions within the cortical midline structure (CMS), regions that do not include the somatosensory cortex. More specifically, as regards somatoparaphrenia, Feinberg et al. (2010), in a detailed study of 13 patients who had been examined by brain imaging techniques within one week of acute hospitalization, identified a pattern of lesions distinctive of those who exhibited its clearest symptoms—repeated, refractory expressions of the conviction that their limbs belonged to someone else. In this study, the region found to be most distinctive was not the somatosensory cortex; rather it was the orbitomedial frontal cortex. The claim here is not that any one region of the brain plays an exclusive, causal role in the etiology of somatoparaphrenia. The claim is that there is no empirical reason to suppose that what underlies the distinctive phenomenology of somatoparaphrenia necessarily involves incapacitation of areas critical to somatosensory processing.

The only reason left to suspect that FB might not actually have experienced the sensations would be the worry that her case is analogous to blindsight. In the case of numbsense (Palliard et al. 1983, Gallace and Spence 2008, and Rossetti et al. 2005)—also referred to as “blindtouch”—although subjects lack conscious tactile experience, they are to some degree capable of non-consciously processing tactile information. In other words, perhaps FB was informationally-sensitive to being touched but not experientially-sensitive to being touched. But FB’s case could not have been an instance of numbsense.

For one thing, well-studied cases of numbsense involve damage to the primary somatosensory areas, very much unlike the case of FB. More importantly, in cases of numbsense the ability to make verbal report is lost. The reason given

for ascribing numbsense is that the patients are able to point, with a moderate degree of accuracy, to the place where they were touched. In other words, by analogy to blindsight, their success at guessing based upon non-conscious information, is indicated by pointing, not by verbal report. FB's case is clearly not like this, since her capacity for verbal report is well-preserved.

I conclude that we are not immune-to-error in the way that IEM indicates. FB's introspections give rise to puzzling responses, responses that are not compatible with IEM. Shoemaker's (1996: 273) critical mistake might have been to infer from "what can happen as a matter of course," to what must necessarily be true of introspection and mental states.

3. Mental states are only contingently related to belongingness: knowing-that a mental-state exists is distinct from knowing-to-whom it belongs:

Zahn et al. (2008) report the case of a 23 year-old male (DP) who complained of "double visions." DP sought medical treatment for this problem five weeks after their acute onset. The "double visions" had begun while he was taking a long-distance flight, during which he experienced tachycardia, shortness of breath, and a fear of asphyxiation.

It was soon established that he does not literally experience double-vision. In fact when looking at a new object he sees it as a single object. But something had changed. According to DP (Zahn et al. 2008, 398), "he was able to see everything normally, but that he did not immediately recognize that he was the one who perceives and that he needed a *second step* to become aware that he himself was the one who perceives the object."

In most other respects DP appeared healthy: for example, the second step was not necessary when initiating actions or when perceiving the actions of others. He seemed not to have passive experiences of his body, changes in body image, memory problems, delusions of control, thought insertions, obsessions, or compulsions. He performed well on a wide range of examinations that included tests for lexical retrieval, for visual object recognition, for attention or executive deficits, and for short-term, working, semantic and episodic memory. Moreover, his medical history contained no indication of psychosocial stress or trauma. Indeed, he seemed socially well-adjusted and capable of managing daily activities. The only other symptom was distress caused by the "double visions." A follow-up exam one year after the initial presentation revealed that the "double visions" continued unabated.

The apparent cause of this condition is hypometabolism, problems pertaining to the supply of or ability to metabolize glucose. In DP hypometabolism was found in right inferior temporal, parieto-occipital and precentral regions. Since “double visions” were restricted to visual object recognition, it is not surprising that the right inferior temporal and parieto-occipital regions were involved, as they are known to be critical to visual object and visuospatial recognition. The former is necessary for the representation of objects as part of the ventral visual “what” stream; the latter, part of the dorsal visual “where” stream.

4. Intimations of the Principles-of-Ownership: Pain Asymbolia, Visual-Tactile Synesthesia, and Empathy

Some states that we know via introspection feel as though they belong to others. Pain asymbolia (Aydede and Guzeldere 2002: 272-275, Lane 2008: 151-153 and Sierra 2009: 150) can help to illustrate this point, for some patients describe their pains thus: “I feel pains in my chest, but they seem to belong to someone else, not to me.” What appears to have happened in such cases is that patients retain the capacity for making sensory discriminations but lack the usual affective responses. The reason the two can dissociate is that sensory discrimination is subserved by a lateral pathway that terminates in the somatosensory cortex, while affect and motivation are subserved by a medial pathway (connected to insular and cingulate cortices as well as to limbic structures). Here, to use Carruther’s (2000: 206) felicitous phrase, the patient “floats above” the pains.

On the other hand, some states that we know of via perception of the external world can, in a qualified sense, feel as though they belong to self. Visual-tactile synesthesia can serve as an example. Synesthesia is a phenomenon wherein the stimulation of one sensory modality evokes the simultaneous subjective experience of sensation in another; perhaps the most common of which is grapheme-color synesthesia, the perceiving of numbers or letters as inherently colored (Robertson and Sagive 2004). But synesthesia takes many forms, including one for which the mere visual perception of another person being touched on the face or neck is experienced as tactile stimulation on one’s own face or neck (Blakemore et al. 2005). One subject claims to have always “perceived observed touch on other people as touch to her own body” (Blakemore et al. 2005: 1573); indeed, she was surprised to learn that the experience of feeling touches applied to other people is not commonplace. In this case the neural mechanism in

virtue of she “felt” the tactile sensations of people she observes in the world seems to be caused by a neural substrate in which her “mirror system”—something which we all have—was overactive, in particular her somatosensory, pre-motor, and anterior insula cortices.

Both pain asymbolia and visual-tactile synesthesia are rare. But other phenomena which help to show how it is we attribute ownership are quotidian, in particular, empathy, our capacity for understanding how others feel. Here we can study, in experimental contexts, what is shared when we know that a mental exists in others and when we know that a mental exists in self. Likewise, we can study what is not shared. For example, Ochsner et al. (2008) compared the direct application of noxious stimuli to the viewing of video clips wherein persons underwent accidental injuries (e.g. leg or arm breaks). They discovered that when self experienced pain, a large portion of the mid insula and a portion of the middle frontal gyrus were uniquely activated. When viewing the video clips of others experiencing pain, the premotor and superior parietal cortex (implicated in shifting attention or perspective-taking) as well as three regions implicated in memory and affective learning (the rostralateral prefrontal cortex, the medial orbitofrontal cortex, and the amygdala) were uniquely activated. But whether noxious stimuli are applied to self or whether one witnesses injury to others, the anterior cingulate cortex and the anterior insula are highly active.

The point of these examples is *not* to illustrate that people are confused about ownership. In the case of pain asymbolia, it just doesn't feel as though it belongs to self. In visual-tactile synaesthesia, subjects know (or can easily infer) that someone else is also feeling a sensation of touch—independently of the tactile sensation that they feel. And, in standard cases of empathy, we are not confused. In an important, but qualified, sense, we do feel what others feel. But we still make appropriate self-other distinctions.

Nevertheless, these phenomena do show three things: (a) pain asymbolia shows that affect and a particular pattern of brain activity might play a role in precipitating a feeling or rendering a judgment as regards ownership. (b) Visual-tactile synaesthesia shows that observation of that which is external to self can be sufficient to induce, if not a numerically identical sensation, at least a sensation of the same type in the observer. And, (c) empathy, precisely because it enables us to study that which overlaps and that which does not, shows that, at least in principle, we could appropriately modulate the non-overlapping

mechanisms so to induce ownership where otherwise it would not obtain. Collectively, these are just three intimations, three clues or points of entry, to teasing apart the recognition of a mental state's existence from the sense of to whom it belongs.

D. Methodology

1. Introduction: collaborations among philosophers and neuroscientists

Ever since the publication of Patricia Churchland's (1987) *Neurophilosophy*, collaboration between philosophers and neuroscientists has become frequent. In some instances these collaborations result in the writing of neuroscience-informed manuscripts as, for example, Patricia Churchland's (e.g. 1998: 231-254) work on Antonio Damasio, or Owen Flanagan's (e.g. 1996: 32-52) work on Alan Hobson. But often the works are more fully collaborative, especially as involves research into the problems of self-consciousness and related subject matter. Representative of these collaborations are Daniel Dennett and Marcel Kinsbourne (1992), Olaf Blanke and Thomas Metzinger (e.g. 2008), Shaun Gallagher and Anthony Marcel (1999), Patricia Churchland and Terrence Sejnowski (1992), Frederique de Vignemont and Patrick Haggard (e.g. Kammers, de Vignemont, and Haggard 2010), Walter Sinnott-Armstrong and Michael Gazzaniga (Sinnott-Armstrong et al. 2008), to name just a few. In these instances philosophical concepts are often employed so as to inform experimental design and clinical investigations, and empirical findings can be so employed as to reshape, revise, and refine philosophical concepts. My focus is on the cultivation of just such fruitful interaction.

國科會，人文處 has explicitly acknowledged the importance of interdisciplinary work of this type by allocating funds for the purchase of fMRI equipment to 國立政治大學, 國立台灣大學, and 國立成功大學. One important intent of this funding is to promote research on issues relevant to traditional philosophical concerns. The nature of self, conscious experience, and the relationship between these is just such a concern: one that has long preoccupied philosophers, yet one that is now, just beginning to become empirically tractable.

2. Promoting a synoptic view—"a beautiful linking of facts":

Sellars (1963: 2) once famously pronounced that “the aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term.” Those neuroscientists and psychologists who are prone to reflection on philosophy of science have often noted that too much experimental work is nothing more than just “a game played by its own rules on an isolated playground” (Wackermann 2006). My hope here is to chart a middle course, to meet in the middle: I am not, at this stage, aiming for the “broadest possible sense” in virtue of which things might be said to hang together. I am, however, aiming to collaborate with cognitive neuroscientists, so to find a middle way—something that anchors the abstractions, without allowing them to be confined to “an isolated playground.” MOT is an effort to show how results from diverse fields of study can “hang together.” One contribution that philosophy can make to cognitive neuroscience is the promotion of a synoptic view, what Wackerman aptly refers to as “a beautiful linking of facts.” If this research program proves successful, it will show how certain pathologies of consciousness, certain illusions, certain dissociative states, certain forms of sleep mentation, etc. are related to one another, how they “hang together.” The next step—three years hence, beginning in 2014—then would be to seek an abstract formulation of the scope that Sellars’ definition is intended to express.

3. Philosophical concerns and prediction:

Not without reservation, but to a considerable degree, I endorse many of Quine’s views. Although I cannot claim to be a Quine scholar, I take it that an important aspect of his philosophical views is that our theories must be grounded in prediction. The point is not of course that the main work of philosophers is, necessarily, to be in the business of producing testable hypotheses. On the contrary, “we believe many things because they fit smoothly by analogy, or they symmetrize and simplify the overall design” (Quine 1995: 256). But, in the same passage, while trying to make it clear that such attempts at fitting things into an overall design are not mere “idle fancy,” Quine (1995: 256) adds that our beliefs generate “every here and there, a hypothesis that can indeed be tested.”

A focus of this three-year project is to seek out some of those aspects of theory—in this case, MOT—which fall into the category of “every here and there.” Much of the motivation, as with the prior work on Shoemaker and Rosenthal, derives from philosophical, largely a priori theorizing. But, wherever possible, I

am looking for opportunities to make explicit links, links that can help inform experimental and clinical science in the production of hypotheses that “can indeed be tested.”

4. A priori assumptions and the choice of methodology:

In his commentary on a project conducted by C. M. Yang, myself, and the Sleep Lab team at NCCU (Yang et al. 2010), Wackermann (2010: 1094) expressed a concern about certain a priori assumptions. Our methodology seems to imply that certain regularities in the phenomenon we were studying (the conscious experience of sleep onset) are universal. Group averaging and across-subject statistical reporting can cause researchers to overlook what he calls “idioversal regularities”. In effect, what might be critically important about the phenomenon is that it is so various; it is achievable in many ways. We (Lane and Yang 2010) agree that this may well be true of sleep onset. And, my provisional view is that it is even more likely to be true of mental ownership; we should anticipate the possibility that it might be variously achieved. Sensitivity to this concern, a concern about certain a priori assumptions, is yet another among the contributions that philosophy can make to cognitive neuroscience. It significantly influences the choice of methodology.

5. Conceptual Analysis: explicating “self”

Although “mental ownership” is not a term-of-art in the philosophy of mind, “self” is. Since “mental ownership” implies “self”, “self” is a term that must be explicated with great care. And since there is a long philosophical tradition of grappling with this difficult notion, it would be foolhardy to ignore the wealth of research that has already been conducted by philosophers.

Naturally though, time is a constraint. Since analysis of the concept of “self” is but one component in the development of this nascent theory and in the promotion of a research program, not all recent philosophical research in this area can be adequately addressed. Therefore, necessarily, during the second year of my project—when “self” will be the focus of philosophical analysis—I must be selective, selective in a way motivated partially by pragmatic concerns. Accordingly I will concentrate on those recent works that focus on “self” and that do so in such a way that they attempt, at least to a limited extent, to engage the empirical sciences. Strawson (2009) will be especially important in this regard since, although his work is motivated almost exclusively by familiar philosophical concerns, he (2009: xv) regards his treatise as “a work of psychology (the more

philosophical division), and many of the claims in it are open to, and I believe deserve, empirical investigation.”

6. Normative Concerns: the fitting together of various levels

Science cannot always just proceed by asking “what is the case?” Often it must ask, “how *should* one proceed?” One area of inquiry for which this normative concern is of special importance concerns the integration of distinct levels-of-description (e.g. the mental and neural). How should we regard the relationship between these two? This is a concern about which philosophers, especially philosophers of science, have, arguably, been more insightful than have most practicing scientists. Whether one should be an eliminativist, an ontic reductionist, a theory reductionist, or whether one should seek some form of interlevel integration is not an arbitrary choice. Each entails different ways of thinking about the phenomenon that one seeks to explain, and matters greatly in determining the likelihood of making progress.

Here I adopt an approach similar to that which Craver (2007) has referred to as the search for a “mosaic level integration.” He attends to mechanism, but not in the pursuit of reduction. He makes it clear that, at least upon his interpretation of the history of neuroscience, progress is most likely to be made when one allows that one should look both “downward” and “upward”. “Downward” at least to the extent that one is seeking lower-level mechanisms for a higher-level phenomenon. (Borrowing a phrase from Kitcher, he [Craver 2007: 259] makes it clear that ignoring the mental would yield nothing but “a world of gory details unfiltered by a higher-level perspective”.) “Upward,” in the sense that one seeks to identify the entities and activities, as well as the properties and their organization, in terms of which a mental phenomenon—like belongingness—is constituted. This up-down pincers maneuver, should strive for “mutual manipulability,” such that one should be able to manipulate the neural level by manipulating the mental level, and vice versa. Moreover, one should aim to show how lower-level events “are organized—spatially, temporally, and actively—with other components” such that the mental events might be realized.

When seeking inter-level integration, one must deal with multiple constraints. One such constraint, one that is of special importance here—since I am dealing with a phenomenon, mental ownership, that was not previously recognized—is the accommodation constraint (Craver 2007: 122-128 and 259-261). As studies carried out at the neural and mental levels co-evolve, some measure of mutual

accommodation will likely be required. The history of neuroscience reveals that motivation for such accommodations can be top-down or bottom-up. One not uncommon consequence is that re-characterization of the explanandum is required. I anticipate that characterization of “belongingness” will require much further tinkering and refinement.

The goal of this research program is, in part, to discover the mechanism(s) of belongingness. To take this as a goal is, in Craver’s term (2007: 266), to provide a “scaffold” for constraints, fully aware that characterization of the phenomenon might evolve. Interlevel explanatory linkages are to be forged by identifying appropriate entities and activities, as well as their organization, and demonstrating their specific relevance to the explanandum—belongingness. Unlike reductionist approaches, here it is taken to be a methodological virtue that probes are made at different levels, because each level carries with it presuppositions that are independent of one another. When they converge in such a way as to shrink the space of plausible mechanisms, precisely because their presuppositions are distinct, our epistemic confidence can, justifiably, increase.

7. Belongingness as explanandum and as explanans:

Of course, a central motivation for this research is that belongingness is a phenomenon that needs to be explained. Accordingly, it is here treated as an explanandum. For too often, and for too long, it has simply been presupposed by scientists and treated through conceptual analysis by philosophers. But I (2009, 2010, and Forthcoming) have shown that such attitudes and approaches are inadequate.

It is in some respects analogous to causation, which tends to be presupposed by scientists, and treated only through conceptual analysis by philosophers. To extend the analogy from philosophical work on causation—which has famously been described as “the cement of the universe” (Mackie 1980)—one might say that investigations into mental ownership are investigations into the “cement” of the mind. I am trying to explain this cement.

But, as the ideas become more mature, especially when we are struck by a highly counter-intuitive phenomena—e.g. the fact that a simple rubber hand can be made to feel as though it is mine—accumulated findings concerning belongingness will be put to use as explanans, that is they will help to explain how belongingness obtains. Explanation often proceeds from “how possibly” questions (1965: 428-432): e.g. how possibly could a healthy person be made to feel that a

rubber hand belongs to them, while their actual hand does not? As MOT becomes more sophisticated, it will help to explain “how actually” (Craver 2007: 112-113) belongingness is realized.

8. An evolving relationship between philosophy and science:

Patricia Churchland (2008: 409) has recently written that “the history of science can be seen as a gradual process whereby speculative philosophy cedes intellectual space to increasingly well-grounded experimental disciplines.” She goes on to proclaim that we are now living in an era during which classical philosophy of mind questions about self and consciousness, which once could only have been addressed through a priori conceptual analysis, are now being addressed by the empirical sciences. She adds that: (a) a priori, conceptual strategies “ran up against a torrent of neuropsychological results that clashed with the ‘truths’ of folk intuition”; (b) “because the data are the data, in place of these alleged ‘truths’ arose empirical questions about brain mechanisms”; and, (c) “the mind turns out to be rather different from how it appears”.

Although MOT was born of an attempt to challenge an alleged conceptual truth, IEM, discovered by a priori methods, and to challenge it by drawing upon the resources of empirical science, we must always be careful not to be excessively hasty or sweeping in our dismissal of classical approaches. The best practicing scientists are sensitive to fine conceptual distinctions, in ways that are evocative of sophisticated philosophical analysis. The data are never just the data; they themselves are the result of concatenations of interpretation, and they in turn lend themselves to re-interpretation when attempts are made to match them to theory. And, how things “appear” cannot be so easily dismissed from the study of mind as from the study of physics, at least because “appearances” are part of that which we seek to understand, part of the explanandum.

I believe that Shoemaker, Rosenthal and most other philosophers are wrong about immunity-to-error. But, even if I am correct, that is not the end of the story. It is, frankly, just the beginning. If we are not immune to error, if belongingness cannot be explained as a tautology, then how is it be explained? I believe this and related issues are just now becoming empirically tractable. Nevertheless, part of the aim of this research program and part of the goal of MOT is to seek refinement of relevant concepts, to give plausible interpretations of data relevance, and to explain the appearance. For all three of these purposes philosophical analysis is indispensable.

9. Promotion of a research program, development of a theory:

To a first approximation, what I am trying to promote is an interdisciplinary research program. Lakatos's (1970 and 1981) ideas partially reflect my views. Of course I do not envision my role as that of a historian of science, one who is trying to assess the status of research programs from the perspective of an outsider. Instead, I am trying to bootstrap a theory—MOT—into existence, by drawing upon the limited resources that are available to me—some the result of philosophical analysis, some that derive from new experimental paradigms, some that are to be found in pathological case studies. I liken this effort to Lakatos for several reasons, not the least of which is that it aspires to his sense of “progressive.”

On this view a theory may be said to be progressing when theoretical growth anticipates empirical growth—that is, if the theory predicts novel facts successfully, it is growing. A goal of MOT is to avoid seeking a level of comfort from which one gives only post hoc explanations of chance discoveries or discoveries produced by advocates of rival theories. If the research program is to grow, the theory must be revised in such a way that it need not depend upon post hoc explanations and that it predicts discoveries of rival theories, without sacrificing the core principles which enabled it to achieve its initial successes.

The mental ownership research program is guided by two leading ideas—the two negative theses and the two positive corollaries of MOT that are given below. These, at least for now, constitute the hard core—a set of commitments that cannot easily be abandoned. The seven conjectures, as they now stand, are treated as part of the protective belt. They are more open to change; if they cease to anticipate novel facts, they might need to be abandoned. Alternatively, their failure might imply that something is wrong with the hard core.

So, in a sense, it can be said that the hard core is not so hard. If a research program proceeds well—if it progresses—its motivating theories may well need to undergo change, changes that reflect development in various stages of the development of a central idea. Our understanding of belongingness may well need to change, just as ideas about atoms and gravity have changed, as the theories in which they were embedded were revised.

One reason for giving equal attention to conceptual analysis and empirical research is that the history of science teaches us that occasional anomalies or awkward empirical facts should not, necessarily prompt abandonment of theory.

Many can perhaps be resolved by means of conceptual analysis—or at least formulated in such a way that they can be more effectively operationalized for use in experimental and clinical contexts. Only when theoretical growth begins to chronically fail to anticipate empirical growth, should the research program be deemed a failure. I believe that by developing MOT in the context of a research program, we are heeding Wackermann’s (2010) wise counsel and are creating the opportunity for scientific progress.

E. Results and Discussion, Part I: Self-Consciousness

i. Self-Consciousness and Immunity

One of the most seminal contributions to the understanding of self-consciousness over the last half century has been Sydney Shoemaker’s articulation of the idea that we are

“immune to error through misidentification relative to the first- person pronouns” (IEM).¹ Along with related ideas developed by Wittgenstein, Castenada, Evans, Perry, and Pryor,² IEM has proven to be extremely fertile in stimulating insights into the first-person per- spective, “the distinctive way mental states present themselves to the subjects whose states they are.”³ Moreover, Shoemaker’s formulation of the idea has motivated significant interdisciplinary research into self-consciousness.⁴

Since first formulating his position, Shoemaker has done much to elaborate upon IEM and related notions. For more than four decades he has been perspicaciously developing his ideas on identification- freedom, introspection, self-knowledge, and the self-intimation of mental states. Although some aspects of Shoemaker’s views on immunity have been disputed, IEM itself has never been severely threatened by any empirical challenge.⁵

Perhaps the most substantial empirical challenge thus far attempted has been Campbell’s⁶ claim that schizophrenic thought insertions, understood in terms of the Frith⁷ monitoring model, might serve as a counterexample to IEM. Gallagher and Coliva have defended IEM by (among other things) arguing that since schizophrenic thoughts are still within a patient’s stream of consciousness, IEM is not vio- lated.⁸ They hold that, as a matter of conceptual truth, “if a subject is introspectively aware

of a certain mental state, then she herself is having it and, therefore, that mental state is her own.”⁹

In this paper we argue that IEM fails. In section i, we adumbrate Shoemaker’s version of IEM along with related concepts central to his understanding of self-consciousness. We also reject the interpretation of IEM as a tautology, and propose to treat it as a hypothesis. In section ii, we present a clinical case—somatoparaphrenia—and in section iii we describe an experiment with healthy subjects—the Body Swap Illusion. In the former case, patients represent experienced sensations as belonging to someone other than self. In the latter, an illusion is created whereby subjects feel that they can shake hands with themselves. The one concerns bodily sensations; the other, the sense of agency.¹⁰ Both cases reveal that IEM lacks modal force: what IEM says cannot happen, can happen. In section iv we respond to possible criticisms of our position. In a concluding section we emphasize that in order to account for the phenomena which seem to defy IEM-based expectations, there is a need to distinguish the ownership of mental states from the ownership of body parts. Moreover, concerning the former, there is a compelling need to distinguish between mental states that are instantiated and mental states that are represented as belonging to oneself.

i. shoemaker’s immunity principle

In his reflections on self-consciousness, Shoemaker takes as a point of departure what he regards as an incontrovertible conceptual truth: “an experiencing is necessarily an experiencing by a subject of experience.”¹¹ He evinces that a subject and an experience are just as intimately related as are a branch and a branch-bending. He then proceeds to develop a conception of self-consciousness that aspires to compatibility with both naturalism and certain Cartesian intuitions.

Developing one among these intuitions, and taking his lead from Wittgenstein, Shoemaker marks a distinction between the use of “I” (and its cognates) “as subject” and its use “as object.”¹² Use-as-subject refers to such expressions as “I am in pain”; use-as-object refers to such expressions as “I am bleeding.” Imagine, for example, that a base-runner and a catcher collide at home plate. As is not uncommon, the catcher’s leg is gashed by the spikes on the base-runner’s shoes, although the catcher does not immediately feel any pain. Because their limbs are entangled, upon first seeing the wound, the catcher does not immediately recognize it as his. As they disentangle, and as the catcher notices distinguishing features like the differences in their uniforms, he comes to realize that it is he who is bleeding. Recognition from the

outside, so to speak, as in identifying the source of the bleeding, is recognition of self-as-object. The experience of pain, by contrast, given that it is known through introspection, typifies knowing about the self “as subject.”

Wittgenstein’s guiding intuition, one which is endorsed by Shoemaker, is: “...there is no question of recognizing a person when I say I have tooth-ache [sic]. To ask ‘are you sure it is you who have pains?’ would be nonsensical.”¹³ Shoemaker identifies the following as prototypical expressions of self-as-subject: “I feel pain”; “I am waving my arm”; and “I see a canary.”¹⁴ Clearly he believes that his argument is applicable to bodily sensation, to the sense of agency, and to perception of the external world. Take “I see a canary,” for example: I might be mistaken concerning what I actually see (it might be a cardinal). I might even

be hallucinating. But “it cannot happen that I am mistaken in saying this because I have misidentified as myself the person I know to see the canary.”¹⁵

Why should it be nonsensical to query whether one is certain that it is oneself who is experiencing the mental state? Because, Shoemaker maintains, when we make a judgment like “I feel pain,” we are aware that “one does, oneself, feel pain...one is, tautologically, aware, not simply that the attribute feel(s) pain is instantiated, but that it is instantiated in oneself.”¹⁶ Accordingly, it simply “cannot happen that I am mistaken in saying ‘I feel pain’ because, although I do know of someone that feels pain, I am mistaken in thinking that person to be myself.”¹⁷ The same is true for judgments about hand-waving or seeing canaries. Notice that these cases exude the modal force of “cannot.” According to Shoemaker this is what makes self-consciousness special.

Shoemaker further elucidates IEM. He says that to claim that a statement “a is F” might be erroneous through misidentification relative to the term “a” is to allow for the following possibility: “the speaker knows some particular thing to be F, but makes the mistake of asserting ‘a is F’ because, and only because, he mistakenly thinks that the thing he knows to be F is what ‘a’ refers to.”¹⁸ But for IEM statements, mistakes of this type are not possible. If the ground of my judgment is introspection,¹⁹ whenever I say “I feel pain” it cannot be the case that I am mistaken in thinking that the person in pain is me. Likewise, whenever I say “I am waving my arm” or “I see a canary” it cannot be the case that I have erroneously identified myself as the person who waves his arm or sees the canary.

How is it that immunity should obtain in such cases? Shoemaker replies that the relevant mental states are identification-free. He believes that even when we need to identify self (as-object), identification “will always presuppose the prior possession of

other first-person information.”²⁰ If self-consciousness always involved identification, whenever

we self-ascribed a mental state (for example, “a is F”) we would need to establish both “b is F” and “a is b.” But “b is F,” in turn, would further require that “c is F” and “b is c” be established. To avoid an infinite regress, we must allow for first-person knowledge that is not grounded in an act of identification.

To illustrate this concern, consider the following. If I notice someone on a shopping center video display, I might wonder whether it is me. In order to make a proper identification, I might pull on my cap while checking to see whether the person on the video display does likewise. To perform this act of identification I must know that I myself am pulling on my cap. How can I know that? According to Shoemaker, my first-person knowledge that I am pulling on my cap must be grounded in identification-free first-person knowledge, because the only alternative would be just the sort of vicious infinite regress schematized above.²¹

Identification-freedom is also integrally related to his views on introspection, the self-intimating character of mental states, and the impossibility of self-blind creatures. For Shoemaker, introspective knowledge refers to just routine, mundane sorts of knowledge.²² In his reflections on how best to understand introspection, he rejects “inner sense” models, notably the “object perception model” (OPM) and the “broad perceptual model” (BPM).

According to Shoemaker, if OPM is correct, then “identification information” about the perceived object must be available.²³ Critically, these objects would need to be independent of acts of perception. But Shoemaker denies that there is any such role for awareness of self-as-object to play in the explanation of introspective knowledge. Although it might appear to be the case that self is a good candidate for being an object of perception, Shoemaker believes that when we do need to identify self-as-object, identification “will always presuppose the prior possession of other first-person information.”²⁴ Again, the only alternative to freedom from identification would be profligate identification, identification that cannot but lead to vicious infinite regress.

Shoemaker also rejects BPM, which differs from OPM in concerning itself with facts rather than objects.²⁵ Despite this difference, though,

BPM shares a fundamental commitment to the view that in perception we have access to things that are independent of being perceived. So identification-freedom would be incompatible with this model too.

Shoemaker's rejection of BPM is also linked to his rejection of the possibility of "introspective self-blindness." He believes that a significant—and unacceptable—consequence of BPM is that it allows for the logical possibility of this particular kind of blindness.²⁶ To be introspectively self-blind with respect to certain kinds of mental phenomena would require that, despite being able to conceive of those phenomena (just as the blind can conceive of phenomena unseen), a creature would be unable to introspectively access them. According to Shoemaker, BPM is only worth taking seriously if self-blindness is regarded as a conceptual possibility.²⁷ But he regards this notion to be as absurd as the claim that we could have pains but be systematically and blithely unaware of them.²⁸

In short, in addition to IEM, Shoemaker endorses a "modest Cartesianism," a "weak version of the self-intimation thesis" (WST). On this view, the existence of certain mental entities is constitutively related to their being available to introspection. For those mental states that have phenomenal character, for example, it is of their essence that having them "issues in the subject's being introspectively aware of that character, or does so if the subject reflects."²⁹ There might well be internal states to which we do not have introspective access, states that play an important role in causing behavior. But Shoemaker says such states would not count as mental. The proper way to think of the relationship between introspection and mental states is that "the reality known and the faculty for knowing it are...made for each other—neither could be what it is without the other."³⁰

Most philosophers regard IEM as a semantic or conceptual thesis. Recall that, according to Shoemaker, when one proclaims self to be in pain "one does, oneself, feel pain...one is, tautologically, aware, not simply that the attribute feel(s) pain is instantiated, but that it is instantiated in oneself." Unlike Shoemaker, we do not regard this as a tautology. On the contrary, it can be subjected to empirical investigation. Our main thesis is: awareness that mental states are instantiated does not entail awareness that said states are instantiated in self. Unlike most critics of Shoemaker, for the sake of argument, we grant

that most of his views are correct. Even so, we argue that genuine counter-examples to IEM exist.

ii. iem and bodily sensations: somatoparaphrenia

Somatoparaphrenia is a syndrome that is characterized by the sense of profound estrangement from parts of one's body.³¹ It is typically found in patients who have suffered extensive right-hemisphere lesions (usually vascular).³² Lesions in the

temporoparietal junction are a common neural correlate of somatoparaphrenia, but deep cortical regions (for example, the posterior insula) and subcortical regions (for example, the basal ganglia) are also sometimes implicated.³³

Somatoparaphrenia is also closely associated with proprioceptive impairment and often (not always) co-morbid with hemispatial neglect. Patients feel that a contralesional limb, most frequently the hand, seems not to belong to them; indeed, it often seems to belong to someone in particular, not uncommonly an acquaintance.³⁴ The sense of disownership can be so vivid that even after recovery patients continue to describe the estrangement in factive, not metaphoric, language.³⁵

Baier and Karnath assessed the frequency of somatoparaphrenia's occurrence.³⁶ They recently examined 79, consecutively admitted, acute stroke patients with right brain damage. They found that 11 experienced estrangement: five exhibited asomatognosia, and six were afflicted with somatoparaphrenia. Of the six, two attributed ownership of the limb to their wives, three to their examining physicians, and one to a patient sharing the same room.

Somatoparaphrenia is occasionally accompanied not only by hemispatial neglect, but also by tactile extinction (the loss of conscious

tactile perception) in the estranged body part. Moro et al. demonstrated (for two cases) that merely by changing the position of the hands—moving the left hand across the midline of the body, over to the right-hand side—tactile sensation could be recovered.³⁷ Even though tactile sensation could be so readily recovered, the sense of limb disownership was unchanged.

As regards our challenge to IEM, the most relevant case has been described by Bottini et al.³⁸ A woman (“FB”) reported that her left hand belonged to her niece and that she, FB, felt no tactile sensations there. In FB's case the lesion was subcortical, involving the basal ganglia, white matter underlying the insula, as well as other areas. But, importantly, the primary somatosensory area—which is critical to processing tactile sensation—was preserved. As Bottini et al. record: “F.B.'s spared ability to perceive tactile stimuli, provided that these were referred to someone else's body, was evidently based on the survival of some elementary somatosensory cortical functions.”³⁹

In a series of controlled tests, FB, while blindfolded, was advised that the examiner would touch her left hand; next, the examiner would in fact touch the dorsal surface of FB's hand. Whenever this was done, FB said that she could feel no tactile sensations. When advised that the examiner was about to touch her niece's hand,

however, upon being touched FB reported feeling tactile sensation. Here we begin to see the relevance of FB's case to IEM.

It should be borne in mind that FB was in all other aspects cognitively sound.⁴⁰ Moreover, in order to ensure that these tests would be reliable, catch trials—wherein FB was led to expect touches that were not forthcoming—were used. These trials were evenly distributed across three verbal warnings—I am going to touch your right hand, your left hand, and your niece's hand—and were administered in four sessions, two on one day, two on the next. It is of paramount importance to note that in not even one of 36 catch trials, nine each per session, did FB respond incorrectly.⁴¹ In other words, when advised that she (or her "niece") would be touched, if no contact was made, FB always reported "no," no contact had been made.

As we begin examining IEM in light of this case, let us assume, for the sake of argument, that Shoemaker's central theses are largely correct. WST is true; both OPM and BPM, false. Moreover, self-as-subject is indeed distinct from self-as-object.

But even if we grant to Shoemaker his principal theses, we are left with an explanatory puzzle: why is it that when FB is expecting to be touched (on the left hand), she feels nothing, whereas when she expects that her niece will be touched there, she is able to report tactile sensation? Why, despite the experimental controls in place (for example, blindfold and catch trials), is she able to judge that "her niece" has been touched? Typically, to say (a) "I am going to touch your arm," implies (b) "I am going to touch you." It would be nonsensical to say (a) without implying (b). Likewise, when the doctor says, "I am going to touch your niece's hand," she implies, "I am going to touch your niece." The concern here is not where the sensation will be felt, but who will feel the sensation. Pace the prototypical situations that motivate the Wittgenstein-Shoemaker intuition—it is not absurd to inquire as to who is the subject of experience.

Let us divide the experiment into two parts: FB expecting that she will be touched is Part 1. FB expecting that her niece will be touched is Part 2. FB's case should be regarded as directly relevant to IEM because she has been primed by the doctor to introspect. We argue that in Part 2 FB is misrepresenting her tactile sensation as belonging to someone else. It is not the case that FB is misrepresenting the location of a sensation, as, for example, the base-runner does if he first represents his own leg as bleeding and then discovers that the bleeding leg is attached to the catcher with whom he collided. Instead, in Part 2, from FB's first-person perspective, when introspecting on that tactile sensation she misrepresents

herself, such that she is not the owner of the sensation. In short, FB commits an error through mis-identification regarding just who is the subject of the sensation.⁴²

To repeat, we can concur with many of Shoemaker's central theses: (1) For every mental state there must be a subject who experiences it. Moreover, for the sake of argument, we can agree with Shoemaker's WST. Thus: (2) Every mental state is in principle available to introspection. And we think Shoemaker would be obliged to concede that FB can only have the experience of a tactile sensation in Part 2 by means of introspection.

Although Shoemaker does not explicitly adopt a position concerning the ownership of sensation, a natural interpretation of his views would be as follows. (1) and (2) conjoined suggest: (3) Every mental state is experienced by the one who is currently introspecting that state.⁴³ The position is made explicit by Coliva,⁴⁴ who takes herself to be "vindicating" Shoemaker's claim that "in being aware that one feels pain one is, tautologically, aware, not simply that the attribute feels pain is instantiated, but that it is instantiated in oneself."⁴⁵

We have formulated (1)–(3) in a way that fully accommodates Shoemaker's views. Our argument is that (1)–(3) do not provide sufficient ground to establish IEM. Proponents of IEM fail to take into account that (1)–(3) do not imply: (4) Every mental state is, from the first-person point of view, represented as experienced by the one who is introspecting the state. It is (4) that is needed for IEM to hold. FB's case is a counter-example to IEM because (4) is not true of those cases for which FB is introspectively aware of tactile sensation in Part 2. Although the attribute feels sensation is instantiated from the first-person point of view, it is not the case that the tactile sensation is instantiated in self. FB does not represent it in that way. The two instantiations are not tautologically linked. For IEM to be true, (4) must hold necessarily. But it does not hold with strict necessity; hence, IEM fails.

It is then empirically possible for a subject, while introspecting a mental state (and thereby knowing that someone is undergoing that state), to be in error with regard to who is experiencing that particular mental state. Admittedly, this is counterintuitive. The Wittgenstein-Shoemaker intuition—it is absurd to inquire of the person who introspects and reports a toothache whether it is indeed that person who has the ache—strikes all of us as correct. But empirical inquiry has ways of upsetting the apple cart: it would by no means be absurd to ask of FB whether it is she who has the tactile sensations, even though it is she who produces the introspectively based report.

An important contrast here calls for explanation. We have a fact and a foil,⁴⁶ the contrast between the two parts of FB's case. In Part 1, when FB is primed to introspect on what she experiences, she reports nothing; in Part 2, when she is primed to introspect on what her niece experiences, she reports tactile sensation.

To ignore this difference would be to ignore a significant explanatory problem. Because Parts 1 and 2 have similar histories, it is possible to ask sensible contrastive questions that enable us to elicit causal differences.⁴⁷ This possibility is not permitted by IEM. The essential difference between the two parts is whether FB represents herself as the subject of the mental state. This issue, concerning first-person representation of just who the subject is, we refer to as mental ownership.

One might worry that FB merely reports feeling the sensation, when in fact she does not feel anything. But had there been no actual sensations on this series of tests, the reports would not have been made. First, recall that in FB's case the lesion was subcortical; her somatosensory cortex was preserved. So it is not surprising that she retained the capacity for experiencing tactile sensations (provided that these were referred to someone else's body). Second, FB's performance on catch trials was perfect. Included among the catch trials were instances in which she was told that her niece's hand was going to be touched, when in fact it was not touched. In these trials she never once made a false report—neither on the first nor the second day of the experiments. Third, other standard procedures were in place to monitor FB's sustained attention and reliability of response: FB's hand was touched in a randomized, fixed sequence, which was repeated in four sessions on two separate days. Because she was blindfolded and because of the other controls that were in place, she could only have given an accurate report had she actually experienced the sensation of being touched.

Moreover, imaging studies of self-referential processing show that there is no reason to suspect that problems pertaining to mental

ownership typically impair tactile processing. Northoff et al.'s analysis of many and varied studies that engage the "feeling of mineness"⁴⁸ indicates that these experiences are subsumed by a set of commonly activated regions within the cortical midline structure (CMS), regions that do not include the somatosensory cortex.⁴⁹ More specifically, as regards somatoparaphrenia, Feinberg et al., in a detailed study of 13 patients who had been examined using brain-imaging techniques within one week of acute hospitalization, identified a pattern of lesions distinctive of those who exhibited its clearest symptoms: repeated, refractory expressions of the conviction

that their limbs belonged to someone else.⁵⁰ In this study, the region found to be most distinctive was not the somatosensory cortex; rather, it was the orbitomedialfrontal cortex.⁵¹ The claim here is not that any one region of the brain plays an exclusive, causal role in the etiology of somatoparaphrenia. The claim is that there is no empirical reason to suppose that what underlies the distinctive phenomenology of somatoparaphrenia necessarily involves incapacitation of areas critical to somatosensory processing.⁵²

The only remaining reason to suspect that FB actually did not experience the sensations is the worry that her case is analogous to blindsight. In the case of numbsense—also referred to as “blindtouch”—although subjects lack conscious tactile experience, they are to some degree capable of nonconsciously processing tactile information.⁵³

In other words, perhaps FB was informationally sensitive to being touched but not experientially sensitive to being touched.

However, FB’s case could not have been an instance of numbsense. For one thing, well-studied cases of numbsense involve damage to the primary somatosensory areas, very much unlike the case of FB. More importantly, in cases of numbsense the ability to make verbal report is lost. The reason given for ascribing numbsense is that the patients are able to point, with a moderate degree of accuracy, to the place where they were touched. In other words, by analogy to blindsight, their success at guessing based on nonconscious information is indicated by pointing, not by verbal report. FB’s case is clearly not like this, since her capacity for verbal report is well preserved.

In conclusion, it seems that we are not immune in the way that IEM indicates. FB’s introspections give rise to puzzling responses that are not compatible with IEM. Shoemaker’s critical mistake might have been to infer from “what can happen as a matter of course,” to what must necessarily be true of introspection and mental states.⁵⁴

iii. iem and the sense of agency: body swap illusion

The case against IEM can be made in multiple ways. In the previous section we dealt with bodily sensations. Here we show that similar problems can arise for the sense of agency concerning mental ownership.

Cognitive neuroscientists have been investigating whether specially designed experiments can induce in healthy subjects certain illusions pertinent to bodily self-consciousness. For example, in the case of the “Rubber Hand Illusion” it has been shown that ordinary people can experience an artificial hand as their own.⁵⁵ In these

experiments, investigators primarily have been interested in the ownership of body parts and various phenomena that involve self-as-object rather than self-as-subject. In at least some of these experimental cases, however, issues relevant to IEM and self-as-subject are implicated. Most noteworthy among these is the “Body Swap Illusion.”

In this case the illusory experience of owning a body that belongs to someone else is induced in healthy subjects. Although some among the neuroscientists who discovered the illusion are concerned only with body ownership, we argue that some of their experiments actually involve ownership of mental states. In the particular case described below, subjects can misrepresent themselves as experiencing someone else’s experiences. After describing the experiment, we explain how it violates IEM.

The Body Swap Illusion was first demonstrated in a series of experiments conducted by Petkova and Ehrsson.⁵⁶ In one setting (their Experiment 5), two persons were involved: experimenter and subject. The experimenter wore a helmet outfitted with two closed-circuit television (CCTV) cameras, which transmit signals to a specific place. By positioning the cameras thus, the scenes they registered presented the experimenter’s viewpoint. Wearing a set of head-mounted displays (HMDs), the subject stood face to face with the experimenter. The subject’s HMDs were connected to the two CCTV cameras on the experimenter’s head such that the images from the CCTV cameras played on the HMDs. The effect of this set-up was that the subject, adopting the experimenter’s perspective, visually perceived himself rather than the experimenter.⁵⁷ The subject could see his own body from the shoulders to slightly above the knees. Both experimenter and subject were instructed to extend their right hands and then take hold, as if to shake. During the course of the experiment the two were instructed to squeeze one another’s hands repeatedly, each time for two minutes. In the illusion condition, they squeezed in a synchronous manner; in the control condition, they squeezed asynchronously, alternating, the experimenter responding to the subject semi-randomly.⁵⁸

Twenty subjects participated in this experiment, and each was interviewed immediately afterwards. The authors claim that the experiment “demonstrated that this set-up evoked a vivid illusion that the experimenter’s arm was the participant’s own arm and that the participants could sense their entire body just behind this arm.”⁵⁹ To obtain more objective, quantifiable data, the scientists incorporated an anxiety-inducing threat into the experimental design (a knife above the wrist to suggest cutting of the hand) and measured each subject’s skin conductance response

(SCR). They reported that they “observed significantly stronger skin conductance responses when the knife was moved near the experimenter’s wrist than when it was moved towards the participant’s own hand in the synchronous condition.”⁶⁰

This experiment has significant implications for IEM. Note that in describing the participants’ phenomenology, the authors say, “after the experiment, several of the participants spontaneously remarked...‘I

was shaking hands with myself!’”⁶¹ Although the subjects “could clearly recognize themselves and distinguish between their own arm and the arm of the experimenter,” this illusion is so robust that “a participant can face his or her biological body and shake hands with it without breaking the illusion.”⁶²

How should this aberrant experience be understood? The most natural way to characterize the subjects’ phenomenology is with respect to agency. When they experience the illusion of shaking hands with themselves, their experiences involve misrepresentation of action awareness—that is, the misrepresentation concerns “who” shakes their hands. This poses a problem for Shoemaker’s IEM.

From the subjects’ first-person point of view, the handshaking experience involves the awareness that I am the agent of this action. This recently has been called “agentive experience” or “agentive self-awareness”—I experience myself as someone who is doing something.⁶³ What is distinctive about the Body Swap Illusion is that the subjects’ agentive experience is mistaken. Although it was really the experimenter who was shaking their hands, the subjects misrepresented themselves as the agent of the action.⁶⁴

To see how this creates a problem for Shoemaker’s IEM as regards the case in question, we can agree with Shoemaker on each of the following: (1) For every agentive experience there must be a subject who experiences it. (2) Every agentive experience is in principle available to introspection. (3) Every agentive experience is experienced by the one who is currently introspecting it. The crucial point, however, is that (1)–(3) together do not imply: (4) Every agentive experience is, from the first-person point of view, correctly represented as experienced by the one who is introspecting it. Without (4) the ground upon which IEM stands is shaken.

Recall that one of Shoemaker’s prototypical examples of self-as-subject is “I am waving my arm.” According to IEM, it cannot happen

that I am mistaken in saying “I am waving my arm” because although I do know of someone that is waving his arm, I am mistaken in thinking that person to be myself. I am necessarily aware that I am, myself, waving my arm. *Mutatis mutandis* for

handshaking. This too clearly involves agentic experience. But as the body swap case shows, having an experience of handshaking does not guarantee that the agentic experience cannot be misrepresented. The mode of representation matters. Here, while in the illusory state, I am introspectively aware of the shaking hands, but I misattribute agency. I commit an error that violates IEM.

In the case of somatoparaphrenia, a subject violates IEM because she experiences a mental state in virtue of having represented that state as belonging to her niece. In the case of body swap, subjects violate IEM because they represent themselves as agents when plainly they are not. In both cases IEM is violated. Introspective awareness that a mental state is instantiated, *pace* Shoemaker, does not prevent us from error as regards mental ownership.

iv. response to possible criticisms

In this section we consider three possible objections. The first concerns the relationship between conscious experience and reportability. The second concerns whether IEM should be regarded as a conceptual truth, and the third concerns an alleged distinction between agency and ownership.

First, the reason FB's case is particularly troubling for IEM is that it consists of two parts which reveal an explanatory contrast. In Part 1, when told that she will be touched, FB does not feel the sensation; yet, in Part 2, when told that her niece will be touched, she feels the sensation. The contrast exhibited here provides strong support for the claim that the self-as-subject of the relevant mental state is misrepresented in Part 2. Since FB felt the tactile sensation in Part 2, why didn't she feel it in Part 1? The only difference between Parts 1 and 2 concerns how the subject, from the first-person perspective, represents with regard to who is to be touched.

To salvage IEM, one might consider an alternative interpretation of her responses. Perhaps FB actually felt the sensation in Part I but, due to her pathologies, just could not report them. Were this so, the critical issue posed by this case might turn on the ability of FB to report tactile phenomenology, not the phenomenology itself. Proponents of IEM then could argue that IEM remains unchallenged because it does not presuppose a necessary connection between reportability and phenomenology. They could argue that FB felt the sensations both in Part 1 and Part 2, so it did not really matter whom the doctor said would be touched. It was just that FB failed to report it in Part 1.

Successful defense of IEM, however, requires that this strategy not remain mere speculation. There must be some reason to suggest that it accurately describes what

transpired in FB's case. But no well-motivated reason suggests itself. On the contrary, there is good reason to think our interpretation of FB's case is accurate. Once again, recall that in order to ensure the reliability of FB's reports, the doctors conducted several catch trials that were evenly distributed across three different prompts: your right hand, your left hand, and your niece's hand. When untouched, FB never reported any sensation. When her right hand was touched, she always, unerringly, reported sensation. When her left hand was touched, she never reported sensation. But when her "niece's" hand was touched, she recovered tactile sensation. There is simply no evidence to suggest that reportability was a problem for her.

A second possible criticism is related to Campbell's interpretation of schizophrenia.⁶⁵ Campbell takes IEM to be a datum in need of explanation.⁶⁶ Indeed, he acknowledges, in accord with Wittgenstein and Shoemaker, that people take for granted the absurdity of asking, "Someone has a headache, but is it me?" Nevertheless, he does suggest that "there is some structure in our ordinary notion of the ownership of a thought which we might not otherwise have suspected."⁶⁷

Coliva criticizes Campbell's interpretation of schizophrenia, and it is her defense of IEM that might abet those who would argue against our position.⁶⁸ Here the concern is just what constitutes mental ownership. We should first emphasize that we do not agree with the entirety of Campbell's argument—neither is IEM a datum, nor is the Wittgenstein-Shoemaker intuition veracious. What we do share with Campbell is the contention that ownership, as it pertains to conscious experience, is more complex than typically is acknowledged.

Responding to Campbell concerning the ownership of mental states, Coliva contends, "If a subject is introspectively aware of pain, this just means that she is feeling pain...it is a matter of conceptual truth that if a subject is introspectively aware of a certain mental state, then she herself is having it and, therefore, that mental state is her own."⁶⁹ She emphasizes that, as a matter of conceptual truth, introspective

awareness of a mental state guarantees that one is the owner of said state. In developing her position she contends that other than introspective awareness, there simply is no independent criterion for what is to count as ownership of a conscious state. She regards this as a "vindication" of Shoemaker's treatment of IEM as a tautology.⁷⁰

But Coliva's view cannot save IEM. To see why, for the sake of argument, let us agree that there really are no independent criteria. The problem then would be that lack of independent criteria would by no means imply that mental ownership cannot

be misrepresented. The key statements (1)–(3), as discussed in the previous sections, can be reformulated as follows: (1) Every mental state belongs to a subject. (2) Every mental state is in principle available to introspection. (3) Every mental state belongs to the one who is currently introspecting that state. Notice that Coliva’s view is fully accommodated by (3). As we have argued above, however, (1)–(3) together do not imply: (4) Every mental state is represented (from the first-person point of view) as belonging to the one who is introspecting the state. Coliva’s objection fails because she neglects (4).

Furthermore, it is not clear that there are no independent criteria for determining mental ownership. Recall what transpires in the case of FB: when advised that she would be touched, she felt nothing. When advised that her niece would be touched, she felt tactile sensations. If we regard IEM as a tautology, if we believe that introspective awareness guarantees mental ownership, then we arbitrarily dismiss the possibility of discovering independent criteria. Such dismissal would be tantamount to begging the question. What has been discovered in the cases of somatoparaphrenia and the Body Swap Illusion is that first-person representation of the ownership of mental states does not comport well with what might seem to be logically necessary or conceptually guaranteed.

FB recovers sensation because she has been cued not to represent the touch as an experience of her own, but as an experience that belongs to her niece. In other words, how the subject represents the experience provides an independent criterion for determining mental ownership. There is no question as to whether or not it is FB who is providing a report based on introspection. So there is no denying that information concerning the tactile sensation is available to FB. But, from the first-person perspective, this is not the end of the story. Ownership of mental states is a more complex phenomenon than the received view of IEM allows.

To the self-as-subject, from the first-person perspective, it matters just how the relationship between self and sensation is represented. Call to mind that one of Shoemaker’s projects has been to elucidate “the distinctive way mental states present themselves to the subjects whose states they are.”⁷¹ What we have found is evidence that the distinctive way mental states present themselves to subjects varies and that, for one form of representation, ownership is contentious. It seems we have cases for which it would be by no means idle or absurd to inquire as to whether the experiences of which a person is introspectively aware belong to that person.

A third possible objection pertains to Gallagher's distinction between agency and ownership.⁷² He employs this distinction to deny that schizophrenic thought insertion causes problems for IEM, because he believes there is no doubt as to just where, experientially, these thoughts are. The patient might well be sincere in expressing the feeling that he is not the author of these thoughts, but that is not to deny that these thoughts occur within his stream of consciousness. "His judgment that it is he who is being subjected to these thoughts is immune to error through misidentification, even if he is completely wrong about who is causing his thoughts."⁷³ In short, although the patient disclaims authorship, he does not deny experiencing the thought. Even in schizophrenia, there remains a nontrivial sense in which the inserted thoughts belong to the patient. Accordingly, in view of the fact that the relevant error in the Body Swap Illusion concerns agentic experience, one might try to argue that Gallagher's distinction applies here as well.

There are several reasons why this defense of Shoemaker's IEM fails. (i) For those subjects who feel that they are shaking hands with themselves during the illusion, it would still be reasonable to ask Wittgenstein-Shoemaker questions: is it you who is having the experience of squeezing your own hand? Is it you who is shaking hands with yourself? Arguably the most compelling intuition that motivates IEM is that questions of this type are absurd. But here they are not absurd. This very fact—that these questions can be well motivated—indicates that IEM does not enjoy the kind of modal force claimed by Shoemaker.

(ii) Admitting misrepresentation of agentic experience is already detrimental to Shoemaker's IEM. Again, one of Shoemaker's prototypical expressions of self-as-subject is "I am waving my arm." As is

the case with "I am in pain," "I am waving my arm" enjoys an absolute immunity. Unlike "I am in pain," though, here we have a clear instance of agency. Shaking hands, just like waving, implies agency. In other words, Shoemaker's elucidation of IEM would still be assailable.

Recall Shoemaker's formal articulation of his claim: for a statement "a is F" to be erroneous through misidentification relative to the term "a" is to allow for the following possibility: "the speaker knows some particular thing to be F, but makes the mistake of asserting 'a is F' because, and only because, he mistakenly thinks that the thing he knows to be F is what 'a' refers to." But for IEM statements, mistakes of this type are not possible: whenever I say, "I am shaking hands," it cannot be the case that I am mistaken in thinking that the person who is shaking my hand is me. It cannot be

the case that I have erroneously identified myself as the person who is shaking my own hand. But that is precisely what happens in the Body Swap Illusion.

(iii) Gallagher argues that the distinction he discerns in schizophrenic thought insertion is sufficient to rescue IEM. But schizophrenia is not analogous to the Body Swap Illusion. First, in the case of body swap the error does not concern a lack of agentive experience; instead, it involves the erroneous attribution of agency to oneself. In the illusory state, one takes credit for more—not less—than one is capable of. Violation of Shoemaker’s IEM in this instance is not due to denial of agency.

Second, in describing the schizophrenic’s attribution of agency, Gallagher observes, “with respect to agency, he is in a position to make only statements in which he uses the first person pronoun as object—and in such cases the immunity principle is not at stake, and therefore cannot be violated.”⁷⁴ In other words, Gallagher’s view is that because schizophrenics lack agentive experience, they can only regard the “author” of inserted thoughts as object. Schizophrenia, then, is conspicuously different from the case of body swap. In the latter case, the subject misattributes agency to self. That is to say, the agent is regarded not as object, but as subject. Since here the agent is a self-as-subject, Gallagher’s distinction cannot safeguard IEM.

(iv) Finally, and decisively, this strategy simply would not work for the case of somatoparaphrenia. No parallel argument drawing upon Gallagher’s distinction can be made. The somatoparaphrenia case involves no action on the subject’s part whatsoever.

In previous sections, we argued that adequate explanation of somatoparaphrenia and the Body Swap Illusion requires recognition

that the ownership of mental states can be misrepresented. In this section, we have responded to what we regard as the strongest defenses of IEM and contend that they are not successful. We therefore conclude that the best explanation of the relevant cases reveals that IEM, rather than being a conceptual truth, is an empirical hypothesis, open to verification or refutation. Indeed, this hypothesis is confronted by substantive counterexamples.⁷⁵

v. conclusion

We have, for the sake of argument, adhered to the distinction between self-as-subject and self-as-object. According to Shoemaker, “absolute” immunity applies to self-attribution of mental states only as regards the former. What we have

discovered is that, even when concerned exclusively with self-as-subject, we are not necessarily immune to error in the way that Shoemaker claims.

According to Shoemaker, introspective awareness that one feels pain tautologically implies both that (a) the attribute feel(s) pain is instantiated and that (b) it is instantiated in oneself. But the cases examined here reveal that (a) and (b) are just contingently connected. It is important to distinguish between those states which are instantiated in someone and those states—of which one is introspectively aware—that are represented as belonging to oneself. Mental states can be introspectively available to a subject without being represented as owned by the subject.

Accordingly, even when considering self-as-subject, we are not immune to error through misidentification relative to the first-person pronoun. Misidentification is possible because we can represent the ownership of mental states variously. Because the ownership of mental states is surprisingly complex there is no guarantee that subjects will not misidentify the subject of experience.

Others who have considered the type of cases treated herein have given most attention to the disownership of body parts. But from the first-person perspective, the question as to whether one owns a body part is distinct from questions concerning the ownership of mental states. For example, in the Body Swap Illusion, subjects are able to distinguish their own arm from the arm of the experimenter. Nevertheless, the illusion that one is shaking one's own hand persists. Ownership of body parts does not necessarily imply ownership of mental states. By allowing for this possibility we are able to account for what otherwise would be a wholly baffling phenomenon: I can recognize the hand extended in front of me as belonging to someone else while simultaneously feeling that I am shaking my own hand.

What does it matter if the ownership of mental states is complex in the ways that we indicate? Although we do not separately develop the issue here, one implication seems to be that first-person mental states are not identification-free in the way that Shoemaker claims. And since identification-freedom is a linchpin for many of Shoemaker's views concerning mental states and introspection, its loss would betoken significant consequences for other aspects of his views on self-consciousness.

Most philosophers agree that the what of conscious experience can be misrepresented, but that the who can be misrepresented continues to strike many as absurd. Shoemaker's IEM is an articulation of this robust intuition, the intuition that is well expressed by Wittgenstein's rhetorical question. But once these intuitions are

clearly articulated, they are better regarded as hypotheses. Otherwise they can be found to arrest growth in understanding. Failure to allow for possible misrepresentation of the subject of experience leads to failure to ask important questions in empirical contexts.

Shoemaker's articulation of IEM as a conceptual truth was an attempt to say what is distinctive about self-consciousness. As we have argued, however, IEM is neither datum, nor tautology, nor conceptual truth. It is a hypothesis. By showing that mental ownership can be misrepresented, we have exposed IEM's vulnerability. Progress in understanding self-consciousness will require further inquiry into the phenomenon of mental ownership.

*We are extremely grateful to Frédérique de Vignemont, Tim Bayne, Thomas Metzinger, Olaf Blanke, Shaun Gallagher, David Rosenthal, Valeria Petkova, several participants in the ASSC 13 Conference (in Berlin), and to this journal's anonymous referees for their helpful comments on previous versions of this manuscript. Work on this manuscript was, in part, funded by National Science Council of Taiwan research grants 97-2410-H-004-154-MY3 and 97-2410-H-002-184-MY3.

¹ Sydney Shoemaker, "Self-Reference and Self-Awareness," this journal, lxxv, 19 (October 3, 1968): 555–67. In previous work he had attempted to make a similar point by referring to certain self-ascriptions that are "noncriterial." See Shoemaker, *Identity, Cause, and Mind* (New York: Oxford, 2003), p. 2.

² See Ludwig Wittgenstein, *The Blue and Brown Books* (New York: Oxford, 1958). Also see Gareth Evans, *The Varieties of Reference* (New York: Oxford, 1982); James Pryor, "Immunity to Error through Misidentification," *Philosophical Topics*, xxvi, 1–2 (1999): 271–304; and additional representative works in Andrew Brook and Richard DeVidi, eds., *Self-Reference and Self-Awareness* (Philadelphia: John Benjamins, 2001).

³ Shoemaker, *The First-Person Perspective and Other Essays* (Cambridge: MIT, 1996).

⁴ See Jose Bermudez et al., eds., *The Body and the Self* (Cambridge: MIT, 1995); Bermudez, *The Paradox of Self-Consciousness* (Cambridge: MIT, 1998); and Masaharu Mizumoto and Masato Ishikawa, "Immunity to Error through Misidentification and the Bodily Illusion Experiment," *Journal of Consciousness Studies*, xii, 7 (2005): 3–19.

⁵ For example: several essays in Brook and DeVidi, eds., op. cit.; Colin McGinn, *The Subjective View* (New York: Oxford, 1983), pp. 45–55; Evans, op. cit., pp. 188–91; and Bermudez, op. cit., pp. 6–8.

⁶ John Campbell, "Immunity to Error through Misidentification and the Meaning of a Referring Term," *Philosophical Topics*, xxvi, 1–2 (1999): 89–104.

⁷ Christopher Frith, *The Cognitive Neuropsychology of Schizophrenia* (Hove, UK: Erlbaum, 1992).

⁸ Shaun Gallagher, "Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error through Misidentification," in Dan Zahavi, ed., *Exploring the Self: Philosophical and Psychological Perspectives on Self-Experience* (Philadelphia: John Benjamins, 2000), pp. 203–39; Annalisa Coliva, "Thought Insertion and Immunity to Error through Misidentification," *Philosophy, Psychiatry and*

Psychology, ix, 1 (2000): 27–34; and Coliva, “On What There Really Is to Our Notion of a Thought,” *Philosophy, Psychiatry and Psychology*, ix, 1 (March 2000): 41–46.

⁹ Coliva, “Thought Insertion,” p. 28.

¹⁰ In this paper “the sense of agency” and “agency” are used interchangeably. Both terms refer to first-person, conscious experience. For recent discussion of “agentive experience” see Tim Bayne, “The Sense of Agency,” in Fiona Macpherson, ed., *The Senses* (New York: Oxford, 2011). Also see below, note 63.

¹¹ Shoemaker, *First-Person Perspective*, p. 10. ¹² Shoemaker, “Self-Reference and Self-Awareness”; Wittgenstein, *op. cit.*, pp. 66–67. ¹³ Shoemaker, “Self-Reference and Self-Awareness,” p. 556. ¹⁴ *Ibid.*, pp. 557–58. These three prototypical cases are all mental states that

Shoemaker elsewhere describes as “weakly self-intimating”: that is to say, it is the nature of these mental states to intimate themselves to their “possessors.” Cf. Shoemaker, *First-Person Perspective*, pp. 50–52. Shoemaker adds that self-reference of this sort is not restricted to first-person pronouns: names and definite descriptions can self-refer in comparable ways. See Shoemaker, “Persons and Their Pasts,” *American Philosophical Quarterly*, vii, 4 (October 1970): 269–85. The relevant discussion appears in footnotes 3 and 5. Also see Shoemaker, *Identity, Cause, and Mind*, p. 10, fn. 4.

¹⁵ Italics added by the authors. Shoemaker distinguishes between absolute and circumstantial immunity. Our concern throughout this paper is exclusively with absolute immunity as regards the self-attribution of mental states.

¹⁶ Shoemaker, “Self-Reference and Self-Awareness,” pp. 563–64.

¹⁷ *Ibid.*, p. 557. Also see Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture I: The Object Perception Model,” *Philosophy and Phenomenological Research*, liv, 2 (June 1994): 249–69, at p. 258; Shoemaker, *First-Person Perspective*, p. 15.

¹⁸ Shoemaker, “Self-Reference and Self-Awareness,” p. 557. Here we are only concerned with present-tense statements, but Shoemaker claims that IEM also holds for certain memory judgments. See Shoemaker, “Persons and Their Pasts.”

¹⁹ Cf. Pryor, “Immunity to Error through Misidentification,” p. 279; and Joel Smith, “Which Immunity to Error?” *Philosophical Studies*, cxxx, 2 (August 2006): 273–83.

²⁰ Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture I,” p. 258.

²¹ Shoemaker, “Self-Reference and Self-Awareness,” p. 561, and *First-Person Perspective*, p. 196.

²² “The knowledge I have in mind is...the humdrum kind of knowledge that is expressed in such remarks as ‘It itches,’ ‘I’m hungry,’ ‘I don’t want to,’ and ‘I’m bored.’” See Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture I,” p. 249.

²³ *Ibid.*, pp. 252–53. ²⁴ *Ibid.*, pp. 254, 258. ²⁵ Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture II: The Broad Perceptual Model,” *Philosophy and Phenomenological Research*, liv, 2 (June 1994): 271–90.

²⁶ *Ibid.*, p. 273. ²⁷ Shoemaker, *First-Person Perspective*, p. 31. ²⁸ Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture II,” p. 275. ²⁹ Shoemaker, “Introspection and Phenomenal Character,” *Philosophical Topics*, xxviii,

2 (2001): 247–73; cf. Shoemaker, *First-Person Perspective*, p. 31. ³⁰ Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture II,” p. 289. Also see p. 275.

³¹ Giuseppe Vallar and Roberta Ronchi, “Somatoparaphrenia: A Body Delusion. A Review of the Neuropsychological Literature,” *Experimental Brain Research*, cxcii, 3 (2009): 533–51.

³² For an exceptional case, see Rogerio Beato et al., “Transitory Somatoparaphrenia Associated with a Left Frontoparietal Meningioma,” *Journal of Neurology*, cclvii, 7 (July 2010): 1208–10. For competing explanations of apparent lateralization, see Gabriella Bottini et al., “Productive Symptoms in Right Brain Damage,” *Current Opinion in Neurology*, xxii, 6 (December 2009): 589–98, at p. 591.

³³ Vallar and Ronchi, *op. cit.*, p. 548.

³⁴ Instances characterized just by the sense of limb disownership are referred to as *asomatognosia*; instances wherein limb ownership is attributed to someone else—some specific person—are referred to as *somatoparaphrenia*. The two are anatomically distinct. See Todd Feinberg et al., “The Neuroanatomy of Asomatognosia and Somatoparaphrenia,” *Journal of Neurology, Neurosurgery and Psychiatry*, lxxxi, 3 (2010): 276–81.

³⁵ Peter Halligan et al., “Unilateral Somatoparaphrenia after Right Hemisphere Stroke: A Case Description,” *Cortex*, xxxi, 1 (March 1995): 173–82.

³⁶ Bernhard Baier and Hans-Otto Karnath, “Tight Link Between Our Sense of Limb Ownership and Self-Awareness of Actions,” *Stroke: Journal of the American Heart Association*, xxxix, 2 (2008): 486–88.

³⁷ Valentine Moro et al., “Changes in Spatial Position of Hands Modify Tactile Extinction but not Disownership of Contralesional Hand in Two Right Brain-Damaged Patients,” *Neurocase*, x, 6 (2004): 437–43.

³⁸ Gabriella Bottini et al., “Feeling Touches in Someone Else’s Hand,” *NeuroReport*, xiii, 11 (2002): 249–52.

³⁹ Italics added by the authors. Bottini et al., *op. cit.*, p. 251.

⁴⁰ According to the clinical report, FB was “fully oriented in time and space and did not show any other sign of mental deterioration on the Mini Mental State Examination (score: 26/31).” See *ibid.*

⁴¹ *Ibid.*, Table 1.

⁴² One might still worry that FB’s error concerns the location, not the subject of experience. For example, one possible characterization of FB’s Part 2 is “I feel the sensation in my niece’s hand.” One could then argue that the subject of experience is not misrepresented. We should not assume, however, that this spare description can do full justice to the perplexing phenomenology. One problem is that it implies that FB, in feeling the sensation, regards “her niece’s hand” from a third-person point of view. Thus, it cannot fully capture the complex pathology. Why? Not only is the clinician baffled by the experimental results, FB too is baffled by her own paradoxical experiences (*ibid.*, p. 251). Both FB’s experience and her relation to “her niece’s hand” are first-personal, introspective. An appropriate characterization, therefore, must also capture the perplexity from the first-person perspective. A more scrupulous reconstruction of this pathological experience would be: “I am introspectively aware of my niece’s sensation.” Under this reconstruction, the subject of experience, or the ownership of sensation, can be misrepresented.

⁴³ Shoemaker does, however, imply this position. See “Self-Reference and Self-Awareness,” pp. 559–60, 565–67. We could articulate (3) such that it even more completely reflects Shoemaker’s preferred mode of expressing his position by adding two clauses: (a) introspective awareness of phenomenal character occurs “if the subject reflects,” and (b) it is just such introspective awareness that enables us to render judgments of the sort under consideration here, such as “I feel pain” (or, as applied to this case, “I feel a tactile sensation”). The relevant passages from Shoemaker are quoted above: the citation for (a) is indicated in fn. 29 and for (b) in fn. 16. But neither (a) nor (b) is critical to our argument.

⁴⁴ Coliva, “Thought Insertion,” pp. 28–29. ⁴⁵ Shoemaker, “Self-Reference and Self-Awareness,” pp. 563–64.

⁴⁶ See Peter Lipton, “Contrastive Explanation,” in David-Hillel Ruben, ed., *Explanation* (New York: Oxford, 1993), pp. 207–27.

⁴⁷ *Ibid.*, pp. 217–19.

⁴⁸ Georg Northoff et al., “Self-Referential Processing in Our Brain—A Meta-Analysis of Imaging Studies on the Self,” *NeuroImage*, xxxi, 1 (May 2006): 440–57. In characterizing the “feeling of mineness” the authors say that “the self we consider here is an experiential self that mediates ownership of experience.” See p. 441.

⁴⁹ Concerning the role of the CMS, Northoff et al. say, “Taken together, our results suggest that self-referential processing is mediated by cortical midline structures....We conclude that self-referential processing in CMS constitutes the core of our self and is critical for elaborating experiential feelings of self, uniting several distinct concepts evident in current neuroscience.” See p. 440; also see pp. 448–49. As regards the relation between CMS and sensory processing, they report: “Our review of neuroimaging studies reveals a set of commonly activated regions, within the extended CMS, during self-related tasks using a diverse set of sensory modalities. Activation in CMS must therefore be considered independent of the sensory mode within which the self-related stimuli were presented. Such sensory independence of neural activity in CMS can be observed in all domains.” See p. 449.

⁵⁰ Feinberg et al., *op. cit.*

⁵¹ “The orbitofrontal lesion was critical in the development of somatoparaphrenia versus simple asomatognosia.” See *ibid.*, pp. 279–80.

⁵² Cf. many of the cases reviewed by Vallar and Ronchi, *op. cit.*, Table 1, pp. 536–37.

⁵³ See Jacques Palliard et al., “Localization without Content: A Tactile Analogue of ‘Blindsight’,” *Archives of Neurology*, xl, 9 (September 1983): 548–51; Alberto Gallace and Charles Spence, “The Cognitive and Neural Correlates of ‘Tactile Consciousness’: A Multisensory Perspective,” *Consciousness and Cognition*, xvii, 1 (2008): 370–407; and Yves Rossetti et al., “Implicit Body Representations in Action,” in Helena De Preester and VeroniekKnockaert, eds., *Body Image and Body Schema* (Philadelphia: John Benjamins, 2005), pp. 111–25.

⁵⁴ Shoemaker, “Self-Knowledge and ‘Inner Sense,’ Lecture II,” p. 273.

⁵⁵ Matthew Botvinick and Jonathan Cohen, “Rubber Hands ‘Feel’ Touch that Eyes See,” *Nature*, ccxcxi (February 19, 1998): 756.

⁵⁶ Valeria Petkova and Henrik Ehrsson, “If I Were You: Perceptual Illusion of Body Swapping,” *PloS One*, iii, 12 (December 2008): 1–9.

⁵⁷ Recall that one of Shoemaker’s three prototypical examples of introspection is “I see a canary.” See “Self-Reference and Self-Awareness,” p. 557.

⁵⁸ Petkova and Ehrsson, *op. cit.*, p. 4, Figure 6. ⁵⁹ *Ibid.*, p. 5. ⁶⁰ *Ibid.*, Figure 7.

⁶¹ *Ibid.*, p. 5. ⁶² *Ibid.*, p. 1. ⁶³ Tim Bayne provides a clear example of agentive experience: “It’s your first day as a waiter/waitress, and you are pouring water into a glass from a jug. As you pour the water, you experience yourself as an agent. You experience yourself as someone who is doing something, rather than someone to whom things are merely happening.” See Bayne, *op. cit.*, p. 3. Also see Bayne and Elisabeth Pacherie, “Narrators and Comparators: The Architecture of Self-Awareness,” *Synthese*, cvix, 3 (December 2007): 475–91.

⁶⁴ Bodily sensations might play a contributing role: light pressure, intensity, duration, and location on the hand, at the points where the hand is squeezed, can all be experienced. But the illusion—“I was shaking hands with myself”—primarily involves action awareness. Also, as we have argued in section ii, IEM can be violated in the case of bodily sensations.

⁶⁵ Campbell, "Schizophrenia, the Space of Reasons, and Thinking as a Motor Process," *The Monist*, lxxxii, 4 (October 1999): 609–25.

⁶⁶ Campbell, "Immunity to Error through Misidentification and the Meaning of a Referring Term," pp. 91–94; Campbell, "The Ownership of Thoughts," *Philosophy, Psychiatry and Psychology*, ix, 1 (March 2002): 35–39.

⁶⁷ Campbell, "Schizophrenia," p. 610. ⁶⁸ Coliva, "Thought Insertion." ⁶⁹ *Ibid.*, pp. 28, 29. Original essay not italicized.

⁷⁰ Shoemaker, "Self-Reference and Self-Awareness," pp. 563–64.

⁷¹ Shoemaker, *First-Person Perspective*.

⁷² Gallagher, *op. cit.*

⁷³ *Ibid.*, p. 231.

⁷⁴ *Ibid.* Italics added by the authors.

⁷⁵ Thus far we have not discussed perception of the external world—the third of Shoemaker's prototypical cases. But, in personal communication, Roland Zahn has described a recent case which suggests that even here IEM can be violated. (Zahn is a Clinical Research Fellow with the Neuroscience and Aphasia Research Unit at the University of Manchester.) The patient was suffering from right inferior temporal hypometabolism, problems pertaining to the supply of or ability to metabolize glucose. Multiple clinical interviews designed to elicit the patient's phenomenology consistently revealed that visual experience required a two-step process: upon first becoming aware of an object it was not immediately obvious that the object was being seen by self. In order to recognize a visual experience as belonging to self it was necessary to take a second step. This step involved relating what was being seen to who was seeing it. Caution is warranted in interpreting this case, for although the condition persisted for at least two months, all information derives from clinical or diagnostic assessments. No experimental protocols were employed. But if Zahn's description is accurate, then Shoemaker's IEM fails for all three of the prototypical cases.

ii. Mental Ownership and Higher-Order Thought

Mental ownership concerns who experiences a mental state. According to David Rosenthal (2005: 342), the proper way to characterize mental ownership is: 'being conscious of a state as present is being conscious of it as belonging to somebody. And being conscious of a state as belonging to somebody other than oneself would plainly not make it a conscious state'. In other words, if a mental state is consciously present to a subject in virtue of a higher-order thought (HOT), then the HOT necessarily represents the subject as the owner of the state. But, we contend, one of the lessons to be learned from pathological states like somatoparaphrenia is that conscious awareness of a mental state does not guarantee first-person ownership. That is to say, conscious presence does not imply mental ownership.

According to Rosenthal's (2005: 4) transitivity principle, mental states are conscious only if one is in some way aware of them. He champions the view that this principle is implemented by HOTs. Succinctly, the HOTs in virtue of which a mental state can become conscious have the content, 'I am in a certain state' (Rosenthal 2005: 343). As he (2005: 343–44) emphasizes, this awareness of a state as present seems 'direct' and 'unmediated'. The notion of self here is minimalist, just a 'raw bearer'. This leaves room for the possibility that one can describe oneself incorrectly. According to Rosenthal's 'battery model' (2005: 345–48), I might misattribute contingent properties (e.g. personal history) to myself. I might, say, believe myself to be Barack Obama. Nevertheless, Rosenthal highlights the point that we are immune to error as regards the raw bearer (Rosenthal 2005: 354–60). According to this version of immunity, the 'Thin Immunity Principle' (TIP), 'when I have a conscious pain, I cannot be wrong about whether it's I who I think is in pain . . . I cannot represent my conscious pain as belonging to someone distinct from me' (2005: 357). HOTs necessarily refer to both the first-order mental state and the owner, who can be none other than self. Conjunction of the battery model and TIP implies that I can describe myself inaccurately, but I cannot represent my conscious mental states as belonging to someone else.

Liang and Lane (2009), however, have argued that empirical evidence can be adduced to refute this claim. Specifically, in the case of a patient (FB) suffering from somatoparaphrenia (a syndrome in which one feels alienated from parts of one's body) accompanied by tactile extinction (in the alien body part), conscious perception was recovered when the patient was advised that somebody other than herself would be touched (Bottini et al. 2002). As the result of a right hemisphere stroke, FB came to believe that her left hand belonged to her niece. In a series of controlled experiments, whenever that hand was touched, FB felt nothing (Part I). She was not mistaken about her identity, was fully oriented in space and time, and evinced no other indications of mental deterioration. But, surprisingly, upon being told that her niece's hand would be touched, FB experienced tactile sensation (Part II).

We suggest that the concept of mental ownership plays a critical role in explaining the dramatic experiential contrast between Parts I and II. It is our contention that FB's case is best explained by distinguishing mental ownership from conscious presence. Even when characterizing FB's case in a way that is maximally consistent with HOT theory, it seems that although a tactile sensation is consciously

present to her in Part II, her HOT does not represent her as the owner. We have argued that this constitutes a counter-example to Rosenthal's view.

Rosenthal (2010) proposes two criticisms of our view. First, he claims 'it's not at all obvious what representing a state as being present to oneself consists in apart from representing the state as belonging to oneself. So it's unclear what their distinction amounts to'. Second, he argues that FB's recovery of tactile sensation can be explained by HOT theory without violating TIP. We begin with the second objection.

Rosenthal (2010) argues that 'There are two kinds of ownership': (a) 'whom a sensation subjectively belongs to', and (b) 'the apparent bodily location of the sensation'. To illustrate this distinction he cites the phenomenon of phantom limb: 'In addition to being aware of bodily sensations as one's own, we are aware of such sensations as having some bodily location; pains, for example, subjectively seem to be in a hand, foot or other body part.' He understands this apparent location as just one among various qualitative aspects of the pain; in the same way that pains can be sharp, dull or throbbing, so too they can seem to be located in the head, the chest or a limb that doesn't exist. On this characterization, (a) is unaffected. Those who experience phantom pain, still experience the pain as their own.

Rosenthal regards FB's case as analogous to phantom limb. He says that because FB is aware of the sensation in a 'spontaneous, unmediated way', it follows that 'she is aware of the sensation as being her own'. It is just that this particular sensation has a subjective location in her niece's rather than in her body. So the idea is that, although (b) is misrepresented, (a) is not. On this view, Liang and Lane fail to recognize subjective bodily location as an alternative and legitimate notion of mental ownership. Accordingly, FB's case can be accommodated by HOT, without violating TIP.

We disagree. First, to claim that spontaneous, unmediated awareness somehow implies that mental ownership can never be misrepresented is to beg the question. It is one thing to say that, in Part II, FB has a HOT that enables her to have spontaneous and unmediated awareness of the tactile sensation. It is something else to say that FB's HOT represents her, from the first-person point of view, as being the owner of that sensation. The inferential leap from premiss to conclusion is substantial: it should not be assumed that subjective spontaneity or apparent absence of mediation guarantees mental ownership. Liang and Lane's objection is precisely that – the two are not necessarily related in this way.

Second, when applying TIP to the case of pain, Rosenthal (2010) argues: ‘No error is possible about whom I am aware of as having the pain because the spontaneous awareness tacitly identifies the bearer of the pain with the bearer of the awareness’. The problem is, again, there is a gap in this argument. It leaves a critical question unanswered – why can’t identification of the bearer of the pain by spontaneous awareness go astray? As Rosenthal has repeatedly emphasized in his writings (e.g. 2002 and 2005), one of the main virtues of his theory of consciousness is that HOTs can misrepresent.² Indeed, HOT theory allows for the possibility of describing mental states that do not even exist. Given that HOTs, *ex hypothesi*, must refer to both a mental state and to the state’s owner, and given that HOTs can be completely wrong about the first order state to which they refer, it is arbitrary to insist that HOTs cannot be wrong about mental ownership. Liang and Lane’s contention is that HOTs can misrepresent not only the content of first-order mental states but also the subject.³ Spontaneous awareness can obtain in the absence of mental ownership.

Third, Rosenthal takes subjective bodily location to be an alternative notion of mental ownership. But this is mistaken. Note that he treats subjective bodily location as ‘an aspect of the qualitative character of bodily sensations’. In other words, where the subject feels the sensation is regarded as part of the content of the sensation, i.e. part of what the subject experiences. For the sake of argument, we can allow that phantom limb might be explainable in these terms, and that ‘we must understand this apparent location as a qualitative aspect of the pain’.⁴ The problem is, if this view is adopted, it would be a mistake to use subjective bodily location to explain somatoparaphrenia. The two cases are not analogous: in phantom limb who feels the pain is not at issue. The qualitative character of bodily sensations is about what the subject experiences, namely the content of first-order mental state, not about who that subject is. Explaining who in terms of what, treating the former as merely derivative from the latter, is to mischaracterize the phenomenological perplexity of mental ownership. Location and belongingness are distinct. In sum, Rosenthal’s objection fails because he has not established subjective bodily location as a legitimate alternative notion of mental ownership.

Recall that Rosenthal’s first objection is that it is unclear what our distinction between conscious presence and mental ownership ‘amounts to’. One way of responding to this worry is by unpacking the distinction in terms of his theory. For the sake of argument we can agree with Rosenthal on the following points: (1) For every mental state there must be a subject. (2) The subject is aware of conscious mental

states in virtue of having suitable HOTs, such that awareness of those mental states seems unmediated and spontaneous. (3) Every conscious mental state is consciously present to the subject. But (1)–(3) do not imply that every mental state is represented, from the first-person point of view, as belonging to the subject, the one who is currently aware of it in a spontaneous, unmediated way. Thus it can be seen that HOT theory itself allows for the possibility that TIP can be violated. And as FB's case shows, when Rosenthal (2005: 357) proclaims that 'one cannot be wrong about whether the individual that seems to be in pain is the very same as the individual for whom that pain is conscious', he is mistaken. No aspect of HOT theory can be enlisted to justify Rosenthal's inference from presence to ownership.

Why is allowing for the presence-ownership distinction so important? Our exchange with Rosenthal is not – and we believe Rosenthal would heartily agree – merely a parochial, philosophical dispute. Wittgenstein (1969: 66–67) once famously claimed that to ask of a person who reports being in pain 'are you sure that it's you who have pains?' would be nonsensical. Most contemporary philosophers have taken this remark to be undeniably true. Wittgenstein, we contend, was wrong. It would not necessarily be nonsensical. On the contrary such questions should sometimes be asked.

Getting clear about the conceptual issues in this vicinity is essential to making progress on a host of challenging empirical issues. One important role for philosophy, which remains underdeveloped, is to elucidate concepts with an eye towards motivating directed, fruitful inquiry, in both clinical and experimental contexts. Consider again FB's recovery from tactile extinction in Part II. Motivated by realization that spontaneous awareness does not guarantee ownership, a clinician might have pursued an additional line of questioning. Adequate investigation of FB's perplexing phenomenal experience would require that she be asked the Wittgenstein question, albeit in slightly recast form, to wit: 'Are you sure it is you who is feeling your niece's sensation?'

Somatoparaphrenia is surprisingly common, some reports (e.g. Baier and Karnath 2008) indicating that it occurs in as many as 8% of acute stroke patients with right brain damage. The presence-ownership distinction espoused here, we suggest, can motivate a research programme that combines well-designed questions and varied stimuli. For example, probes similar to those employed in the Cambridge Depersonalization Scale (Sierra and Berrios 2000) could help to evince and render reportable the rich phenomenological complexity. Along with this scale, multifarious

stimuli should be applied. The tactile tests on FB can be supplemented with, for instance, the cold pressor pain test (e.g. Mitchell et al. 2004), aiming for a more refined, nuanced understanding of her phenomenology.

Making salient the distinction between mental ownership and conscious presence, and wielding these notions perspicaciously, is a significant way in which philosophy can contribute to the development of revelatory empirical inquiry. To illustrate with a current dispute among neuroscientists, Feinberg et al. (2010) have recently criticized the Geschwind–Gazzaniga account of somatoparaphrenia as incapable of explaining the ‘bizarre aspects of the confabulations displayed by our somatoparaphrenia patients’. He proposes an alternative account that has clear implications for distinguishing between the neuranatomical substrates of asomatognosia and somatoparaphrenia. Proper evaluation of these competing empirical accounts, we submit, requires that serious attention be given to those ‘bizarre aspects’, most notably the phenomenology of mental ownership.

1 The order of authorship was determined arbitrarily; this article is completely collaborative.

2 Rosenthal (2005: 8) touts this as a clear advantage that his implementation of the transitivity principle has over rival implementations, like inner-sense models.

3 Lane and Liang (2009) have previously shown that Rosenthal’s battery model of self-identification does not prevent TIP from being violated, at least as regards the case of FB.

4 Although we do not argue the point here, some theories, like Melzack’s (1989) ‘neuro-matrix’, suggest that Rosenthal’s approach might even fail to adequately account for phantom limb phenomena (both pain and other sensations).

iii. Higher-Order Thought and Pathological Self

(1) Introduction

Somatoparaphrenia, a pathology of self, is philosophically perplexing. It poses a significant challenge for theories of consciousness, including David Rosenthal’s higher-order thought (HOT) theory, which holds that HOTs are scientific posits in a theory that aims for explanatory adequacy. In a recent series of papers Rosenthal (e.g. 2005: 341) has employed the HOT theory as part of an attempt to explain ‘our sense of having a unified consciousness’, a ‘sense’ which he understands as the ‘compelling intuition’ that we have a single self. He develops his explanation in terms of an immunity-to-error principle (thin immunity), which holds that we are immune to error in certain restricted judgements concerning self. After presenting Rosenthal’s theory

in x2, in x3 we argue that it fails to explain somatoparaphrenia, a pathology in which mental states can be conscious even when they are represented as belonging to someone other than self. We discuss some possible responses in x4 and, finally, in x5, we point out a broader implication of this empirical challenge to the HOT theory.

(2) HOT, self and the thin immunity principle

According to Rosenthal's HOT hypothesis (e.g. 2002a: 408–11), a mental state is conscious just in case it is accompanied by a suitable, first-person thought to the effect that one is in that state. First-order mental states become conscious only if they are intentionally targeted by thoughts that are occurrent, assertoric, and seemingly non-inferential, thoughts which can represent the state as being present. Importantly, on this view, to represent a state as being present just is to represent it as belonging to somebody (Rosenthal 2005: 342). So a HOT in virtue of which a first-order state becomes conscious must both refer to that state and to the owner of that state (Rosenthal 2004: 160–61). Simply put, HOTs have the content, 'I am in a certain state.' This reference to I, understood as the owner of the state, is 'unavoidable' (Rosenthal 2005: 342, 347). It follows from this necessity claim that 'being conscious of a state as belonging to someone other than oneself would plainly not make it a conscious state' (Rosenthal 2005: 342).

Self, as characterized by the HOT theory, is minimalist (Rosenthal 1997: 86): it is a 'raw bearer' in that nothing about the way it is characterized by aHOT distinguishes it from any other self (Rosenthal 2005: 342–45). The raw characterizations of self provided by HOTs do not enable self-identification. Identifying oneself consists of saying who one's first person thoughts are about, and this identification is accomplished by reference to a diverse 'battery' of contingent properties, properties that include matters of personal history, bodily and psychological characteristics, and current circumstance (2004: 212, 2005: 345–48). Appropriately, Rosenthal refers to this as the 'battery model' of self-identification. These descriptive identifications of the self can be erroneous: it is empirically possible, for example, that I take myself to have the contingent properties of Barack Obama.

Although we can self-identify ourselves erroneously, Rosenthal (2004: 168–76, 2005: 353–60) believes that we are immune to a certain type of error of misidentification. He (2005: 354–60) refers to this as 'thin immunity' to indicate a contrast with Shoemaker's (1968: 557) stronger concept of immunity. As it applies to body sensations, Rosenthal (2005: 357) says of the Thin Immunity Principle (TIP)

that, 'when I have a conscious pain, I cannot be wrong about whether it's I who I think is in pain.' And why is this? The reason is to be found in the very idea of HOT. According to Rosenthal (2005: 346), HOTs are first-person thoughts; and, for example, my pain state's being conscious consists in my being conscious of myself as being in pain. It follows then that 'I cannot represent my conscious pain as belonging to someone distinct from me' (Rosenthal 2005: 357).

This form of immunity is thin in the sense that it is consistent with the battery model, for I can still be wrong about just what contingent properties I possess. I can, for example, believe that I possess the properties possessed by Barack Obama, as opposed to those that are actually mine (Rosenthal 2004: 177–78). In developing this idea, Rosenthal proclaims that when I look at myself in a mirror, I can be wrong in many ways; I can extravagantly mis-attribute properties to myself, thinking that I am Obama. To do so would not constitute a violation of TIP. But what Rosenthal (2005: 359) insists upon is that, 'if I think I see myself in a mirror, I cannot be wrong about who it is I think the individual in the mirror is.'²

And why might we be immune to error in these ways? Although Rosenthal nowhere states the point explicitly, TIP is a direct consequence of the HOT theory. Recall, according to the theory, a mental state is conscious just in case it is accompanied by a suitable HOT such that one is conscious of oneself as being in that state. Because every HOT is a first-person thought, it has a unique owner and it necessarily represents its owner as the unique rawbearer of first-order sensory states. It follows then that we are thinly immune to these errors concerning bodily sensations or visual perceptions.

Rosenthal believes that the HOT theory and TIP can accommodate both quotidian and pathological states, including more than just misidentifications of the sort already mentioned. Concerning a hypothetical Dissociative Identity Disorder (DID) case wherein a patient appears to have two selves, Rosenthal (2002b: 215–6) says: First, DID cases are patients with partially disjoint sets of first-order mental states. Although the sets partially overlap, coherence tends to be higher within than between them. Second, Rosenthal posits disjoint sets of HOTs, each targeting distinct portions of the partially disjoint first-order states. Third, he proposes that the apparent sense of two distinct selves can be explained by the battery model, because the patient employs partially disjoint sets of contingent properties to identify the individual who the first-person thoughts are about. By appealing to disjoint sets of first-order states,

disjoint sets of HOTs, and the battery model, Rosenthal argues that the appearance of distinct selves is explainable by the HOT theory and that TIP is not violated.

(3) Violation of the thin immunity principle by a pathological self

We have argued that Rosenthal's TIP is implied by the HOT theory. If this is the case, violation of TIP would constitute a serious problem for the theory. Below we argue that there is indeed empirical support for the claim that TIP is sometimes violated.

Somatoparaphrenia (Vallar and Ronchi 2009) is a syndrome that is characterized by the sense of alienation from parts of one's body. It is typically found in patients who have suffered extensive right-hemisphere lesions (usually vascular), but it can also be caused by subcortical lesions (for example, in the basal ganglia). Patients typically feel that a contralesional limb belongs to someone other than self. Baier and Karnath (2008) examined 79 acute stroke patients with right brain damage and found that six were afflicted with somatoparaphrenia. Of the six, two attributed ownership of the limb to their wives, three to their examining physicians, and one to a patient sharing the same room.

This syndrome is frequently accompanied by the loss of conscious tactile perception in the alien body part. Bottini et al. (2002) describe the case of a woman (FB) who reported that her left hand belonged to her niece and that she (FB) felt no tactile sensations there. In a series of controlled tests, FB, while blindfolded, was advised that the examiner would touch her left hand; next the examiner would in fact touch the dorsal surface of FB's hand. Whenever this was done, FB said that she could feel no tactile sensations. When advised that the examiner was about to touch her niece's hand, however, upon actually being touched, she reported feeling tactile sensation. To monitor attention in and the reliability of these tests, catch trials were distributed across three verbal warnings – I'm going to touch your right hand, your left hand, and your niece's hand – were administered in four sessions, two on one day, two on the next.

If we describe this case in the terminology of the HOT theory, what seems to be happening is that these tactile sensations are represented as belonging to someone other than self. That these states can be conscious seems to consist in FB being conscious of her niece as being touched. But if this is so, then we have a clear violation of TIP. Recall that according to Rosenthal HOTs are first-person thoughts that both represent mental states and represent self as the owner of those states. According to TIP, which is derived from this core idea, it should be the case that FB

represents the sensations as belonging to herself. HOT and TIP do not allow for the possibility that the sensations could be represented as belonging to FB's niece. But the empirical evidence presented above confounds this theory-based expectation.

Notice that FB is not failing to identify herself correctly. Unlike the sort of pathological case that the HOT theory is allegedly able to handle, FB is not misidentifying herself as her niece. FB is not attributing a battery of her niece's contingent properties to herself. Rather she is representing herself as not being the raw bearer of the tactile sensations. So Rosenthal's battery model of self-identification cannot be invoked to help preserve TIP.

We contend that this pathological case shows that TIP is sometimes violated and that allowing for the violation of TIP enhances our understanding of the phenomenological aspect of mental states. To insist on TIP would be to risk obscuring a significant empirical phenomenon. Allowing for violations of TIP, given that it derives from the core ideas of the HOT theory, creates doubts about the theory itself.

(4) Possible defences of HOT and TIP

First, one might insist on trying to explain the case of somatoparaphrenia along the lines of that which Rosenthal suggested for the hypothetical Dissociative Identity Disorder case. Perhaps, it might be suggested, there are independent sets of HOTs that target only partially overlapping first-order states, HOTs that give rise to independent personalities. But unlike DID, here the analogue of a DID alter, the niece, does not have a distinct personality that is able to take control of the body and make first-person reports. So there are no grounds for arguing that the subject has multiple, independent sets of HOTs that serve as the foundation for distinct persons.³

A second objection might be that subjects' first-person reports are confused and thereby unreliable. After all subjects are reporting experiences. And when viewed through the lens of the HOT theory these experiences simply could not be reported were they not represented as belonging to the subject who reports them. Any descriptions to the contrary, especially those that are produced by victims of pathology, should be dismissed.

But dismissal of patient reports in these cases would be much too quick. HOTs, by hypothesis, are posits of an empirical theory, and a main reason given for believing they exist is that they are reportable (Rosenthal 2005: 313–14). To be reportable and accurate lends more support to an existence claim than does to be reportable but massively erroneous. So Rosenthal should tread lightly here. To simply

dismiss these (and other) perplexing reports would be to risk ignoring a phenomenon that requires explanation.

Theories that aspire to enhance their empirical credentials do not progress by ignoring anomalous explananda. And the most natural reading of the somatoparaphrenia case is that TIP is violated. Were Rosenthal to insist that TIP holds and that subject reports pertaining to the ownership of mental states are completely in error, he would need to assume the burden of at least showing how these anomalous reports can be accommodated by the HOT theory. And to be successful in this endeavour it would not be sufficient to merely posit disjointed mental states and the battery model, for in the previous section that strategy has already been shown to be inadequate.⁴

A third possible line of objection would be to take the reports seriously, but to reinterpret them. One might, for example, resist a literal understanding of them. Perhaps when FB reports on her niece's tactile sensations there is a sense in which FB might still be the actual owner, even though the way it seems to her causes her to misattribute the ownership of the mental states.

But to reinterpret the case of somatoparaphrenia in this way would be inconsistent with Rosenthal's explanatory intentions. Recall that Rosenthal's intent is to explain the 'sense' or 'compelling intuition' that we have a single self. He is not talking about the physical realization of mental states or about an actual self. So any attempt to distinguish between how things are and how things seem would be inconsistent with the goals of TIP and the HOT theory. What matters just is the appearance, that compelling intuition. Currently though the best evidence we have concerning the proper characterization of appearance in the case of somatoparaphrenia – subject reports – suggests the conclusion that TIP can be violated. And the violation of TIP in turn suggests what appears to be a fundamental problem with the HOT theory: it does not allow for a distinction between the representation of a mental state as present for someone and the representation of a state as belonging to someone. But it is far from obvious that theoretical considerations should be allowed to trump the available empirical evidence, especially given that there is indeed conceptual space between presence and belonging.

(5) Conclusion

The focus of our attention here has been Rosenthal's HOT theory and TIP. We argue that certain pathological phenomena are best explained by allowing that TIP does not always hold. And to allow that TIP does not always hold is to raise serious

questions about the presuppositions upon which the HOT theory is grounded. But our conclusions have implications for other theories of consciousness as well. Consider, for example, Kriegel's (2005) claim that phenomenal consciousness necessarily involves both a (i) what- it-is-like aspect and a (ii) for-me aspect. If the conclusions reached here are correct, Kriegel's views and the views of others who posit a necessary connection between (i) and (ii) are wrong. Just as there is significant conceptual space between presence and belonging, so too is there significant conceptual space between what-it-is-like and for-me.⁵

1. The order of authorship was determined arbitrarily; this manuscript is completely collaborative.

2. Below our argument focuses on the version of TIP which concerns body sensations, but we suspect that the perceptual (the mirror) version might also be susceptible to empirical challenge. Cases of mirrored-self misidentification (e.g. Breen et al. 2000 and Postal 2005) raise the possibility that even if I think I see myself in a mirror, I can be wrong about who it is I think the individual in the mirror is.

3. Although we do not argue the point here, we suspect that the phenomenon of intra- consciousness, wherein one alter claims to be aware of the mental states of other alters (e.g. Wilkes 1993: 112–27), suggests that TIP might not even accommodate all of the experiences that occur within DID.

4. Rosenthal (2005: 209–13) does allow for the possibility of HOTs that misrepresent, even HOTs that completely misrepresent the content of first-order mental states (Lane and Liang 2008). But TIP strictly prohibits misrepresentations concerning the ownership of mental states by the raw bearer of those states.

iv. A soft self and a hard core.

Introduction: Andy Clark's claims that (i) the mind extends into the body and world, that (ii) the boundaries of the body are fluid, and that (iii) we are designed so as to seek out opportunities for mind extension, might all be true. But his defense of these claims relies somewhat excessively upon his misunderstanding of pathological cases like Alzheimer's Disease (AD). Worries pertaining to the role that AD plays in his arguments lessen somewhat the support for (i)-(iii). More significantly though, these worries expose substantial difficulties with his attempts to explain self as the result of soft assembly. Moreover, these worries seem to lend some degree of support to a claim (made by Bruce Sterling) that Clark forcefully repudiates: we might be headed toward a world wherein our peripheral tools are clever but our foundation is weak, vulnerable, pitifully limited, and possibly even senile. I show how Clark can, in principle, by developing the right sort of tools, minimize this worry. Finally though,

I argue that even were we to develop these tools it is unlikely that we could avoid the fate predicted by Sterling.

(1) Some of Andy Clark's claims about the relationships among brain-body-world: A. A claim about where minds can be found:

Extended Mind Thesis (EMT)—cognitive processes can extend into the body and the external world. The focus is on vehicles, not content. B. A claim about our bodies:

Our bodies are negotiable in that we “are essentially open to episodes of deep and transformative restructuring in which new equipment (both physical and ‘mental’) can become quite literally incorporated into the thinking and acting systems that we identify as our minds and bodies” (Clark 2008, 31). A common example is just fluency in using tools, like walking sticks; we come to feel that we are touching the world at the end of the stick rather than touching the stick with our hands.

C. A claim about our brains: *We were shaped by evolution to be neural opportunists*—natural-born searchers for ways to extend our minds. Our intelligence—e.g. capacity for abstract thought—is made possible by NBC.

D. A claim about human nature: Humans (and, in limited ways, all primates) are Natural-Born Cyborg (NBC).

“It is our basic *human nature* to annex, exploit, and incorporate nonbiological stuff deep into our mental profiles” (2003, 6). We are *promiscuous* body-and-world exploiters. We constantly test and explore possibilities for incorporating new resources deeply into problem-solving routines. “This (fact about us) matters philosophically because it invites us to take our best present and future technologies seriously as *quite literally* helping to *constitute* who and what we are” (Clark 2007b, 278).

(2) Some of Clark's claims about the self:

Although Clark devotes most of his attention to specific cognitive performances, still he does sometimes attend to “persisting cognitive agents”—selves. He proclaims the human self to be a *soft* self, “a constantly negotiable collection of resources easily able to straddle and criss-cross the boundaries between biology and artifact” (Clark 2007b: 28). Those resources include the neural, the bodily, and the technological.

The selves that result are called “soft” to reflect the claim that they are the product of “soft assembly” (e.g. Clark 2004: 179), i.e. that they *just are* the transient bindings of heterogeneous, distributed elements into agent-like coalitions.

Clark (2007a, 104-105) is clearly aware that “self” has different referents, but he believes that his approach can account both for the core sense, e.g. having a point of view and a sense of spatial location, and the more complex (perhaps uniquely human) sense wherein we construct narratives concerning (inter alia) what projects and qualities we value as well as what trajectory our lives have taken and what trajectory we hope they will take. Clark (2003: 138-142) denies that there is any central cognitive essence that could be called a self. Not even the narrative self counts.

We literally are just soft selves, transient coalitions that come together to solve problems.

(3) A telling example of what Clark regards as the wrong way to think about us and our futures: Bruce Sterling on “brain augmentation”: “Japan (for example) has a rapidly growing elderly population and a serious shortage of caretakers. So Japanese roboticists...envision walking wheelchairs and mobile arms that manipulate and fetch. But there’s ethical hell at the interfaces. The peripherals may be dizzingly clever gizmos...but the CPU is a human being: old, weak, vulnerable, pitifully limited, possibly senile.” (Sterling is cited by Clark in multiple places: e.g. 2007b: 264 and 277-278; 2008, 30.)

(4) Clark believes: such fears as those expressed by Sterling are shaped by a misguided view of what we already are. Among other things, we are not CPUs trapped in feeble shells; we are soft selves whose “boundaries and components are forever negotiable, and for whom body, thinking and sensing are woven flexibly

(and repeatedly) from the whole cloth of situated, intentional action.” (Clark 2007b: 275).

(5) Emboldened by this view of what we already are, Clark (2003: 139-142) embraces the “cognitive rehabilitation” of Alzheimer’s patients. Specifically he is deeply impressed by the ability of such patients who, despite performing dismally on standard psychological tests, nonetheless cope well with the demands of daily life, because their home environments are “wonderfully calibrated to support and scaffold these biological brains. The homes were stuffed full of cognitive props, tools, and aids.” Examples include: message centers where notes are stored, photos of family and friends with indications of names and relationships, labels and pictures on doors, memory books to record new events, meetings, and plans; and “open storage” strategies of just leaving commonly used things out in the open.

(6) He warns against viewing these people as “hopelessly cognitively compromised” by inviting us to recall just how dependent we are upon pens, paper,

notebooks, alarm clocks, and so forth. The scaffolding in our homes makes it seem that “in a certain sense” a nontrivial sense, our brains are Alzheimic too. The way we scaffold our worlds makes it seem, “in a certain sense,” that we are “exactly” like them.

(7) Alzheimer’s Disease (AD) has played a significant illustrative role in Clark’s views ever since he proposed EMT (Clark 1997):

A. Recall: According to EMT: mental states (e.g. states of believing *p*) can be partially realized by structures outside the head (Clark 2008, 76-82). For example, external traces (e.g. pencil marks in a notebook), under the proper conditions, should (the claim is normative) be regarded as among the physical vehicles whereby some dispositional beliefs are realized. “Proper conditions” are best understood thus: the pencil traces are poised for action in ways that are relevantly similar to internal memory traces. External traces can become so deeply integrated into online strategies of reasoning and recall as to be only arbitrarily distinguishable from the rest of the cognitive engine.

B. The Parity Principle (2008, 77): “If, as we confront some task, a part of the world functions as a process which, were it to go in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process.” In order to identify the physical realizers of cognitive states and processes, we should ignore metabolic boundaries of skin and skull. Instead, we should attend to the computational and functional organization of the problem-solving whole. The parity principle provides a “veil of ignorance” test for helping to avoid biochauvinistic prejudice.

EMT (2008, 87-89) arguments are grounded in “commonsense functionalism” (CSF) concerning mental states. According to CSF “normal human agents already command a rich (albeit largely implicit) theory of the coarse functional roles distinctive of mental states” (Clark 2008, 88). CSF is not equivalent to an empirical functionalism that might only use folk psychology as a starting point for scientific investigation. Moreover, EMT concerns only a subset of mental states that are recognized by CSF, i.e. certain non-conscious, dispositional states such as believing *p*.

D. Alzheimer’s example comparing 蔡 and 馬: 蔡 hears of a demonstration at 晶華酒店. She thinks, recalls that it is on 中山北路, and sets off. 馬 suffers from a mild form of Alzheimer’s, and as a result, he always carries a thick notebook. When 馬 learns useful new information, he always writes it in the notebook. 馬 hears of the

demonstration at 晶華酒店, retrieves the address from his notebook, and sets off.

According to EMT: Just like 蔡, 馬 walked to 晶華酒店 because he *wanted* to go to the demonstration and he *believed—even before consulting his notebook*—that it was on 中山北路. The functional poise of the stored information in the 蔡 and 馬 cases is sufficiently similar to warrant similarity of treatment. The only difference is that 馬's long-term beliefs aren't all in his head.

E. EMT is an active form of externalism. In the 蔡 and 馬 case, the relevant external features are *active*. They play a causal role in the generation of action. If he does not have the notebook traces, then 馬 does not go to 中山北路. If Ma's enemy, No-Neck Bear, tampers with the notebook such that it indicates the demonstration is to be held at 中央研究院, then 馬 goes there instead. Accordingly, “the causally active physical organization that yields the target behavior seems to be smeared across the biological organism and the world” (Clark 2008, 79).

F. A common criticism the 蔡-馬 case (Clark 2008, 80): All 馬 believes (in advance) is that the address is in notebook. That belief leads to his checking the notebook, which in turn leads to his believe about the actual address. Clark's response to this criticism: 馬 “is so accustomed to using the notebook that he accesses it automatically when bio-memory fails.” Checking the notebook is deeply and subpersonally integrated into his problem-solving routines...“The notebook has become transparent equipment for 馬, just as biological memory is for 蔡.”

(8) What is pre-clinical or mild AD really like? Recall (a) Clark's basic claims, (b) his enthusiasm for cognitive rehabilitation, and (c) his disdain for Sterling's claim that we will be left with CPUs that are old, weak, vulnerable, pitifully limited, possibly senile.

EMT: a main reason for believing it is that AD patients treat notebooks as they would biological memory. But in fact this does not seem to be the case with these “peripheral brains” in cases of AD. What one believes—in a biologically-based way—is that information is contained in the notebook. (Black 2001, 49)

Our bodies are open to the world, e.g. the walking stick interface. But for those who use walking sticks, in the early stages of AD one of the first symptoms is a de-emphasis on use, whether for probing deep nooks and crannies (its paradigmatic use in this context), or for balance, lower back support, and so forth (Black 2001, 24-25).

We are neural opportunists, constantly searching for resources that can be exploited into problem-solving routines. But in fact perseveration—a kind of

functional fixedness in problem-solving—begins to set in. No matter how unsuccessful a strategy is, the patient perseveres with it (Ridderinkhof et al. 2002). A marked reduction in cognitive flexibility sets in.

Arguments for EMT are grounded in common-sense functionalism: these lines of argument are highly compatible with the Occupational Therapy approach to AD advocated by Clark (e.g. Baum and Edwards 2003 and Baum et al. 2000). Both are third-person perspectives on the subject. The occupational therapist who hopes to keep the patient in a home environment is closely attuned to the needs of the caregiver, not necessarily attempting to understand the inner life of the patient. Zombie, giant look-up table, Chinese nation, and other similar thought experiments that are commonly employed in criticisms of functionalism in ordinary cases carry even more weight here, for here we have excellent reason to believe that internal states of the patient are seriously compromised. Clark (1993, 214-219) formerly endorsed constraints on the kinds of beings for whom we might properly attribute cognitive processes, but he seems to have loosened his requirements in these regards.

One referent of “self” is having a point of view and a sense of orientation in space. When we read Clark’s antiseptic story about 蔡-馬, 馬 seems to be well oriented in space. But note that AD’s earliest symptom often just is the experience of losing spatial memory, and it is not something that is typically taken lightly (Black 2001, 3-4): “As he rose to begin the morning ritual of preparation, his familiar world changed utterly. He could not remember how to get to the bathroom. He couldn’t navigate a route that he had negotiated thousands of times. After the disbelief and then the denial that anything was amiss, he panicked. Then, suddenly, the clouds parted...(Later that day he went to a florist.) On leaving the store, before any conscious awareness, he was gripped by the forgotten, frightening condition of the morning. Though only two blocks from home, on a route he had traveled for ten years, he didn’t know which was to turn...(The clerk showed him the way.) (He) finally stood in front of his apartment door with a bewildered sense of confusion and relief...” Recall—this is just a telling of the first episode.

(9) How should we regard the seeming tension between Clark’s version of AD and the real thing?

A. It might be claimed that the AD case is not critical to Clark’s application of the parity principle in the defense of EMT. But without the cognitive impairments of AD it would be harder for Clark to motivate his claim that the notebook becomes deeply and sub-personally integrated in problem-solving routines, at least not in a way

that is transparent. When the person in question is not afflicted by AD, or some similar pathology, the claim that some of the active cognitive processes are external is harder to establish.

B. Clark might want to consider an alternative example, e.g. a young person who has suffered severe damage to the hippocampus, thereby interfering with the ability to lay down new memories for the future. Because unlike AD patients some of these amnesiacs suffer from no other cognitive deficiencies, perhaps they could better exemplify EMT. But the actual use of notebooks exhibited by these people is not nearly so fluid or fluent as that which is depicted by Clark.

C. I am not concerned to say that vehicle externalism is wrong. The fact that those who suffer from AD or amnesia in the year 2008 do not have access to external vehicles that can accommodate cognitive processes does not imply that those who so suffer in the future will not have access to such vehicles. I see no principled reason for ruling out this possibility. Indeed I believe that much of Clark's complaining about "bio-chauvenism" is little more than a red herring. For those of us who believe mental states and processes must supervene on physical bases of the right sort, the demand is only for external vehicles of the right sort. In that spirit I do think it important to say that at least some of Clark's examples are falling short of the mark.

The examples in question don't provide external vehicles of the right sort. One reason why this matters is that we should hope that our understanding of afflictions like AD and amnesia are accurate.

D. I suspect that one reason why some people are impressed by Clark's AD example is that people who do so suffer exhibit great effort in trying to maintain a semblance of a normal life. The efforts they make and the efforts of caretakers lead us to give them the benefit of the doubt. We are willing to believe that they have successfully integrated these vehicles into their active mental life.

E. But now imagine two people with memory deficits: (i) one is 90 and suffers from AD. The (ii) other is 19 and suffers from having grown up in a society where various modern instruments coupled with excessive coddling made it possible for him to fail to cultivate the discipline necessary for holding things in mind. In other words, one is sick and one is bone lazy. (I am here excluding the possibility that the 19 year old would be diagnosed as having ADHD and treated with ritalin or adderal for such a diagnosis and pharmacological treatment would not benefit Clark's case.) I suspect that were we to apply the parity principle here, few would be willing to give the 19 year old the benefit of the doubt: that is, few would be willing to say that the external

vehicles have been successfully integrated into an active mental life. Instead, they might question whether an active mental life is here to be found, without regard for external-internal distinctions. They might then be led to worry a bit that something of importance has been left out of Clark's account.

(10) Does Clark commit a mistake similar to that committed by metaphysical behaviorists?

Recall that according to Clark selves are "soft." The self is just a constantly negotiable collection of internal and external resources that come together in transient bindings. For these transient bindings of resources there is nothing that counts as a central cognitive essence. This might be setting a misleadingly low standard for what is to be counted as mentality or agency.

But in his defense, it must be said that Clark is sensitive to various tensions in his account and he does seek to "find a balance" (2007a, 115-118). He notes that a part of mind, a conscious part, acts as an ecological controller—something capable of "adding crucial nudges to the complex dynamics of much larger...systems." (2007a, 115-116) But he also insists that we are soft selves—distributed, de-centralized, self-organizing, and so forth—who just happen to have a perspective on our own activity and a story to tell. Here then he seems to be trying to find a role for the narrator view of self—a view that includes narratives about the projects and qualities we value along with a narrative concerning the trajectory of our lives, both as lived to date and as hoped for the future.

In trying to diminish this tension Clark (2007a, 116) employ's Vellman's (2000, 35-52 and 209-211) notion of "self-fulfilling assertions." Vellman observes that a statement like "I'm going out for a walk" can sometimes be a cause of subsequent walk taking. Prior to uttering the statement the speaker's motivation to walk might not have been sufficient to outweigh countervailing motivations. But by being sufficient to produce the statement, motivations in favor of walking make it more likely that a walk will be taken, because now an additional motive has been added—e.g. "the desire not to have spoken falsely" (Vellman 2000, 209). In effect assertion of the statement raises the price of inaction and seems to be the "crucial nudge" that Clark refers to.

On Clark's (2007a, 116) spin the trick of adding a "narrative-induced cost" to inaction can be effective whether or not it is spoken aloud. Inner rehearsal would be sufficient. The unspoken thought or the overt speech will increase the likelihood of

action because “*the drive for consistency and alignment* acts as a causal influence on what we then do.” (Italics are mine.)

Much more would need to be said here, but I’ll settle for just the question, what is it about Clark’s soft-assembly view of self that entitles him to invoke “drives for consistency and alignment.”?

I am struck by a similarity to behaviorists who upon realizing that rats who were neither rewarded nor punished when maneuvering through a maze nevertheless learned quite a bit about the maze, remarked that the rats exhibited curiosity. Now this is precisely the type of mental state that a self-respecting behaviorist should seek to avoid. Likewise, an advocate of the view that there is no “central cognitive essence” had best avoid talk of “drives for consistency and alignment.”

(11) Some (e.g. Juarro 2004) have pressed similar concerns in reaction to Clark’s work. The worry, as Clark (2004, 179) expresses it, if self is just a constantly negotiable collection of resources that come together in transient bindings, then what “holds it all together?” Just what is or what does the binding?

A. Clark’s (2004, 179) initial response is that “the commonsense ideas of persons, selves, agents, and moral responsibility are all (deeply inter-animated) forensic notions.” In other words the application of these concepts is more a matter of habit and convenience than metaphysical necessity.

B. His (2004, 179) second response is that the process of soft-assembly binding into “agent-like” coalitions will eventually be scientifically tractable. He recommends the “dynamic core” notion developed by Edelman and Tononi (2000) as an approach which suggests how such ideas will become tractable.

C. Although Clark says little about the dynamic core hypothesis, it is easy to see why he is attracted to it. Although the neural substrate of the core is localized in the brain, the core is nevertheless spread out widely, over the entire thalamo-cortical system (Edelman and Toloni 2000, 111-154). (Speaking only very roughly, the thalamus is a kind of relay station for the cortex, such that everything going in or out of the cortex must pass through it.) And the hypothesis itself is an informational or functional hypothesis, one which emphasizes synchronization and coordination of activities in different areas (“re-entry”) along with the mutual sharing of information among those areas (“complexity”). Consistent with Clark’s view of soft assembly, the dynamic core is a process that is defined in terms of interactions whose composition is constantly changing, one set of interactions persisting for no more than a few milliseconds (Edelman and Toloni 2000, 144).

D. But the dynamic core hypothesis has nothing to say about the generation of self-fulfilling assertions, the drive for consistency and alignment, or conscious nudges.

(12) Perhaps Clark should consider yielding on the idea that there is no central cognitive essence. A. Clark is fond of saying things like “the mental buck stops” nowhere and “it is just tools all the way down” (2007a, 111). B. But he typically follows such claims with qualifiers like “some elements...must be more important to our *sense* of self and identity than others. And some elements will play larger roles in control and decision-making than others” (2007a, 112).

C. While he seems to be making a concession in passages of this sort, he follows it immediately with the reassertion that the various neural circuits and external vehicles each make a contribution to “our sense of who we are, where we are, of what we can do, and to decision-making and choice. But no single tool among this complex is intrinsically thoughtful, ultimately in full control, or plausibly identified as the inner ‘seat of the self.’ We (we human individuals) just *are* these shifting coalitions...of tools. We are ‘soft selves,’ continuously open to change and driven to leak through the confines of skin and skull, annexing more and more non-biological elements as aspects of the machinery of mind itself” (2007a, 112).

D. What Clark seems to be doing is this: he doesn’t want to be seen denying the obvious—i.e. that some neural substrates (or external vehicles) and some functions matter more to self than do others. But he then constructs a straw man—i.e. no single tool is the seat of the self. The straw man is then followed by a reassertion of the transience, the fluidness, and the extendedness of these soft selves. The result of this rhetorical sleight of hand seems to be that because most would not want to be associated with the idea of a seat of the self, they unwittingly buy into the idea of thoroughly soft selves.

E. I believe that Clark stumbles into this position, perhaps because he feels that exorbitantly high levels of transience, fluidness, and extendedness are necessary for him to defend EMT and NBC. If we admit talk of CPUs, executive control, and so forth, then it may well be harder for him to make his case that external vehicles are mental and that the brain is designed to aggressively extend itself into the world.

If we allow for a CPU, it might strike some as though the “real” work of mind is still confined to skin and skull.

(13) There is no CPU in the brain but cognitive and neuro-scientific models of mind do utilize concepts like “executive control” and “system override” in

empirically responsible ways (Stanovich 2004, 46). A. Following Stanovich (2004) I'll speak not of a CPU, but of an analytic system which has the capacity for exercising executive control and system override. B. Analytic processes (2004, 44-47) are characterized by serial processing, central executive control, conscious awareness, capacity-demanding operations, and domain-generalty in recruiting information for computation. They allow us to sustain context-free mechanisms of logical thought, inference, abstraction, and planning. The necessary de-contextualizing is expensive and difficult to maintain. For example, knowledge of probability theory, logic, theory-based knowledge, and ethical precepts can enable us to override intuitive, natural responses. C. It seems to be the case that when we decouple—marking mental states as hypothetical rather than actual states of the world—we rely largely upon the analytic system (Stanovich 2004, 50-52). Decoupling enables us to sufficiently distance ourselves from representations of the world that they can be reflected on, even improved. Only in this way can we exercise the capacity for executive control and system override.

D. Clark's emphasis on more natural responses, less expensive cognitive processes, and a fluid exchange between the internal and external causes him to diminish the significance of analytic systems and decoupling. This in turn leads him to adopt a dismissive attitude to Sterling's worries.

E. Let me reassert that I am not advocating bio-chauvenism. For the most part, I think this is a red herring. Vehicle externalism is not so radical a thesis.

Were, for example, the functions of those parts of the frontal lobes that are critical to executive control and override—note that social maturity as codified in law coincides with the completion of frontal lobe maturation (Goldber and Bougakov 2007, 364)—transferable to external vehicles, then arguments on behalf of EMT would carry more weight. The worry then is not that people described by Sterling would be without CPUs; it is that they would lack the capacity for executive control and override, because currently irreplaceable neural resources such as those in the frontal cortex have been compromised. As this relates to AD, even in its very early stages, arterial spin labeling reveals diminished perfusion at several places within the frontal lobes (Ramsay et al. 2007, 502).

(14) Would better technology solve the problem? Were we, say, to (i) develop vehicles that could simulate the functions of the anterior cingulate cortex as well as (ii) develop the means for assuring portable, broad-bandwidth, reliable linkage to the human brain, would the Sterling worry be substantially diminished? A. Clark (2007b,

278-279) is an optimist: because human enhancement is not new, because the conscious mind is comfortable with relying upon external vehicles, and because we can diligently demand that technological prostheses better serve and promote human flourishing, he thinks we have good reason to be cautiously optimistic.

B. But Sterelny (2004) notes that the external vehicles are located in public, often contested, space. Hence they are more directly exposed to subterfuge than is the brain (e.g. recall No-neck Bear's tampering with 馬's notebook). Because of this the external vehicles might have the opposite of the effect intended by Clark: instead of making us more intelligent, we would have to be more intelligent (e.g. more vigilant) in order to use them well.

C. Another problem concerns the unique possibilities of breakdown: the frontal lobes seem to provide an important neural substrate for executive control and override in part just because they are richly connected to other parts of the brain. But the richness of connectivity carries a downside—lesions in other parts of the brain will necessarily have an impact on it (Goldberg and Bougakov 2007, 355). Consequently, the frontal lobes are uniquely fragile. Were their functions to be externalized, breakdown possibilities would more likely increase than decrease.

D. Evidence that we are already becoming less intelligent, not more: chimpanzees outperform us on working memory of numerals (Matsuzawa 2007).

Might this just be a trade-off? We make more space in the brain for the neural substrates which incline us toward EMT and NBC. Perhaps, but the case could just as easily be made that reliance on external vehicles meant that less effort needed to be expended on memory and that there was no selective disadvantage to expending less effort. To make the point somewhat more vivid, consider that by recent standards John von Neuman was a genius. Two aspects of his genius were hypermnesia (i.e. a photographic memory) and the ability to, almost effortlessly, divide two eight-digit numbers in his head (Poundstone 1992, 32-35). Perhaps were it not for the development of external vehicles modern society would be populated by more people like von Neuman.

E. Indeed there might be good reason to believe that we are becoming ever less intelligent. Succinctly, external vehicles make it possible for more of us to

survive than would have been possible had those vehicles not existed. We might need to pay a price for doing what Muller (1997) has described as “relaxing natural selection.” Muller argues that advances in technology, living standards, and medicine have been and may continue to relax the genetic burden that was in effect

under primitive conditions when deleterious mutations greatly reduced survival chances. The problem is that we are passing down to an indefinite number of future generations the burdens that we are spared because we are treated by medical and technological advances. Future generations can also be treated, “but each successive generation will have not only the mutant genes which have been passed along to it but also its own new crop” (Muller 1997, 342). If ameliorative measures succeed, unless they are accompanied by artificial selection, they will lead to an decrease in the reproductive elimination of mutant genes. If we assume that this trend continues indefinitely, the manifestation of mutant genes will continue to rise in total frequency.

It will be necessary for people to reduce the amount of time and energy spent on dealing with the external environment and turn increasingly inward. People “would be devoted chiefly to the effort to live carefully, to spare and to prop up their own feebleness, to soothe their inner disharmonies and...to doctor themselves as effectively as possible. For everyone would be an invalid...” (Muller 1997, 343).

Muller does not apply his argument to the technologies that Clark treats, but it is easy to see how it would apply. To the extent that deleterious mutations that bear on cognitive activity are ameliorated by new technologies, to that extent, we might be just increasing our supply of mutant genes in a way that creates evolutionary drag.

Clark once said that external vehicles enable us to be “dumb in peace.” It might just be that we are becoming more dumb.

F. Results and Discussion, Part II: Sleep Mentation

i. The threshold of wakefulness, the experience of control, and theory

development.

We are very grateful to Professor Wackermann for his constructive and insightful comments. We take it that one among his main concerns is the possibility of variation that might go undetected due to our choice of methodology. For example, when inquiring as to the logicity or coherence of thoughts, we seem to be

presupposing a shared internal norm that can be accurately reflected, despite multiple, mediating steps of memory and evaluation.

Professor Wackermann's concern is by no means an idle one. But (1) we believe that queries of the sort employed here are necessary if we are to begin making progress in eliciting the structure of conscious experience. Were we to limit ourselves to questions that concern just raw, sensory experience, we would risk arbitrarily ignoring significant aspects of the subjects' phenomenology. Further, (2) our statistical analysis is indeed intended to balance out individual differences. Although this strategy does risk obscuring important variation, it can still be helpful in identifying significant, albeit not universal, indicators. Finally, (3) while it is possible that the transition from wakefulness to sleep follows different paths, the issue is an empirical one. Just as it would be unwise to arbitrarily ignore individual variation, so too would it be unwise to arbitrarily inflate individual variation.

We realize that Professor Wackermann's concerns though are not mere methodological quibbles as regards how best to address a single psychological phenomenon. As Wackermann (2006) lucidly expresses elsewhere, he seeks to develop a strategy for discovering universal laws that is compatible with the study of entities that exhibit great variation, human beings. Indeed, we are in sympathy with his view that more attention should be given to what he terms the "idiomatic" regularities. On this view, research should proceed in a two-step fashion: first, one should attend to intra-individual regularities and render these in logical or mathematical form. Only after this step has been completed should one seek inter-individual comparisons.

As regards the research that actuated Professor Wackermann's critique, we are not able to present results in such a way that they would satisfy strict standards for "distributed nomothesis." But, motivated by this strategy, we have re-evaluated the data, attending more carefully to individual variation. In so doing we found that, for most subjects, ratings on more than one item were associated with the perception of falling asleep. Moreover, for ten of the twenty subjects, "control over thinking process" was associated with the perception of falling asleep; for eight subjects, "control over perception"; and, for seven, "thinking experience," "logic of thinking process," and "orientation."

This reanalysis suggests that the experience of control might be a key factor in the subjective experience of sleep onset. Not only is it cited explicitly with reference to thought process and perception, it seems to be implied by those who indicate that

their thoughts were not logical. Speculating, perhaps further investigations of the relevant cognitive processes would reveal that one among the significant idiomatic regularities related to this transitional state involves diminution of the sense of control.

Presumably the relevant meaning of control here does not concern the obsessive or intrusive thoughts that are commonly associated with insomnia—after all, the reported experiences are regarded by the subjects as indications of sleep onset. It would seem to be far more likely that the relevant sense of control bears greater similarity to the thought insertions experienced by schizophrenics, what Frith (1992) refers to as “passivity experiences.” Frith’s account might also help to explain the association with “control over perception,” as his model emphasizes our capacity to distinguish between changes in our perception of the external world that result from our own actions and changes that result from alterations in the external environment itself. In schizophrenics this ability is impaired, an impairment that Frith attributes to a failure to monitor intentions. Inability to monitor intentions might well be experienced as an inability to control perception.

Naturally we do not intend to suggest that sleep onset and schizophrenia are one and the same. Clearly the two differ in many respects. But exploration of the nature and degree of difference might well lead to significant insights.

Such exploratory work would, we believe, be consistent with Wackermann’s (2006) view that science should be dedicated to the search for a “beautiful linking of facts.” He believes that too much experimental work is nothing more than “a game played by its own rules on an isolated playground.” He advocates regarding experimental work as “materialized reasoning”: that is, experiments should be motivated by careful theory development that is relatively independent of particular databases. One goal of such development should be a “beautiful linking of facts,” where previously there had only been a disconnected jumble.

As a very preliminary step in the direction of finding pattern amid jumble, recent research into control of action and goal maintenance suggests a separate, but arguably relevant domain. For example, Suhler and Churchland (2009) have proposed a neurobiological model of control that is applicable both to quotidian states wherein control is exercised (e.g. getting out of a warm bed on a cold morning) and to prototypical cases wherein persons feel “out of control” (e.g. addiction). They propose a model of multiple parameters that includes neurochemicals, connectivity among brain structures, and so forth. The intent is to identify an “in control” region

within multi-dimensional space, a space that reflects the likelihood that there are many different ways of being in, or out, of control.

Were we to further develop such a model, because the prefrontal cortex is implicated in self-monitoring and in goal-directed thought and because it is relatively inactive during NREM sleep (Muzur, Pace-Schott, & Hobson, 2002), we would likely include it as one among several parameters that needs to be highlighted. Nevertheless, we are keenly aware that our brief discussion here merely gestures in the direction of one possible line of inquiry. But an especially attractive feature of such theorizing is that it allows for the possibility of mathematical modeling in a way that can accommodate “idiomatic” regularities: that is, it can account for different ways of being in and out of control. Of course whether or not thinking along the lines adumbrated here will yield fruitful results, we cannot say. But we are grateful for Professor Wackermann’s gentle encouragement to search for the idiomatic and to take seriously the role of theorizing.

ii. What subjective experiences determine the perception of falling asleep during

sleep onset period.1. Introduction

Abstract

Sleep onset is associated with marked changes in behavioral, physiological, and subjective phenomena. In daily life though subjective experience is the main criterion in terms of which we identify it. But very few studies have focused on these experiences. This study seeks to identify the subjective variables that reflect sleep onset. Twenty young subjects took an afternoon nap in the laboratory while polysomnographic recordings were made. They were awakened four times in order to assess subjective experiences that correlate with the (1) appearance of slow eye movement, (2) initiation of stage 1 sleep, (3) initiation of stage 2 sleep, and (4) 5 min after the start of stage 2 sleep. A logistic regression identified control over and logic of thought as the two variables that predict the perception of having fallen asleep. For sleep perception, these two variables accurately classified 91.7% of the cases; for the waking state, 84.1%.

Ó 2010 Elsevier Inc. All rights reserved.

Sleep onset period is defined as the transition from relaxed, drowsy wakefulness to unresponsive sleep. It has been observed that this period is associated with marked changes in a host of physiological and behavioral phenomena, as well as in subjective experience (Ogilvie & Wilkinson, 1984). Physiological phenomena associated with sleep onset include: decrease in high frequency electroencephalographic (EEG) activities (e.g., Azekawa, Sei, & Morita, 1990; Davis, Davis, Loomis, Harvey, & Hobart, 1937, 1938; Hori, 1985; Merica, Fortune, & Gaillard, 1991; Rechtschaffen & Kales, 1968; Tsuno et al., 2002); the absence and presence of different event-related potential (ERP) components (for review, see Campbell, Bell, & Bastien, 1992; Harsh, Voss, Hull, Schrepfer, & Badia, 1994); the appearance of slow eye movements (e.g., De Gennaro, Ferrara, Ferlazzo, & Bertini, 2000; Porte, 2004); the absence of elicited skin conductance responses (e.g., Johnson, 1970); a drop in the core body temperature and an increase in the distal skin temperature (e.g., Barrett, Lack, & Morris, 1993; Krauchi, Cajochen, Werth, & Wirz-Justice, 2000; Wehr, 1990); and, substantial, rapid reduction in respiration (e.g., Colrain, Trinder, Fraser, & Wilson, 1987; Naifeh & Kamiya, 1981). Behavioral indicators of sleep onset include: a decrease in sensory threshold, a cessation of responses to external stimuli (e.g., Anliker, 1966; Ogilvie & Simons, 1992; Ogilvie, Simons, Kuderian, MacDonald, & Rustenburg, 1991; Ogilvie & Wilkinson, 1984, 1988; Ogilvie, Wilkinson, & Allison, 1989; Simon & Emmons, 1956), and a decrease in muscle strength (e.g., Jacobson, Kales, Lehmann, & Hoedemaker, 1964; Litchman, 1974) were also observed in the course of the sleep onset process. And, as regards the subjective experience of sleep onset, loss of awareness of environmental stimuli and the loss of control over thought processes have both been reported (e.g., Foulkes & Vogel, 1965; Gibson, Perry, Redington, & Kamiya, 1982).

Although these different phenomena are all associated with sleep onset, they are not always synchronized. Thus, the criteria for sleep onset identified for different studies are not consistent with one another. Most studies used physiological indices: for example, one of the most commonly used standards for sleep onset – the beginning of stage 1 sleep – is defined as the first 30-s epoch in which EEG alpha activities decrease to less than 50% (Rechtschaffen & Kales, 1968). Other studies, however, demonstrated that the subjective perception of falling asleep was more closely associated with stage 2 sleep, which is characterized by diminished responsivity to external stimuli. Webb (1980) reported that from 66.7% to 85% of those who were physically roused from sleep while in stage 2 sleep perceived this as

awakening from sleep; the others did not feel as though they had been asleep. Even higher rates of discrepancy between physical arousal and the subjective perception of awakening were reported when assessments were made at the onset of stage 2 sleep: in several studies the percentage of those who felt as though they had been asleep was below 50% (Amrhein & Schulz, 2000; Hori, Hayashi, & Morikawa, 1994; Sewitch, 1984).

Naturally physiological definitions of sleep onset lend themselves to stricter methodological controls. Thus, they tend to be accepted as the standard indices of sleep onset. It is commonly assumed that physiological indications of sleep highly correlate with subjectively experienced sleep. Discrepancies between the two tend to be regarded as “sleep-state misperception.” Regarding this as a “misperception” clearly implies that physiological measures are given greater weight. In daily life, by contrast, subjective perception is the most frequently used criterion for sleep onset. People typically judge the amount of time taken to fall asleep merely on the basis of our subjective experience, without the evidence of any objective indices. Relatively few studies, however, have focused on the subjective experiences that reflect sleep onset. Although a few previous studies have examined the correspondence between subjective experience and electrophysiological phenomenon during sleep onset, to the best of our knowledge, no study has explored the subjective experience that determines the perception of sleep onset.

Previous studies revealed that subjective experiences occurring during sleep onset include changes in thoughts, images, or sensations (for review, see Schacter, 1976). For example, Foulkes and Vogel (1965) collected 212 reports on the subjective experience of sleep onset from nine, young and healthy subjects, at four distinct junctures: continuous alpha EEG with rapid eye movements (REMs), discontinuous alpha EEG with slow eye movements (SEMs), descent into stage 1 sleep, and 0.5–2.5 min of stage 2 sleep. The aspects of subjective experience analyzed included sensory imagery, affect, thought control, and reality orientation. Foulkes and Vogel reported that sensory experiences were primarily visual, and remained so throughout the sleep onset period. Thought control and reality testing were found to decrease continuously during the process of falling asleep. Affective experience was minimal to begin with and then decreased even more after one had fallen asleep. Foulkes and Vogel concluded that hypnagogic experience – defined as a state of intermediate consciousness that precedes sleep – was quite similar to REM dream experience. This same pattern of changes in subjective experience was confirmed in other

studies by the same research group (Vogel, Barrowclough, & Giesler, 1972; Vogel, Foulkes, & Trosman, 1966). Similarly, Gibson and his colleagues investigated subjective experiences associated with judgments that corresponded to physiological sleep states during the sleep onset period. They discovered that three cognitive criteria were significantly correlated with correct estimation of physiological sleep states; these three were thought control, awareness of surroundings, and temporal awareness (Gibson et al., 1982). More recently, a study utilized the absence of eyelid and head movements to define sleep onset and collected over 1000 reports of subjective experience from 11 subjects sleeping at home. The results were similar to those of previous studies in that they evinced a decrease in “wake-like” thoughts and an increase in dream-like mentations from 15 s to 5 min following sleep onset (Rowley, Stickgold, & Hobson, 1998).

Furthermore, other studies used a more data-driven approach to identify the cluster of subjective experiences that are associated with physiologically defined sleep onset. For example, a study used canonical correlations to investigate the correspondences between EEG states and subjective experiences, with subjects lying in bed during their typical bedtime. Results showed that peak power in 2–6 Hz and 13–15 Hz bands as well as low power in 9–11 Hz and 16–25 Hz bands were associated with a cluster of subjective experiences. These experiences included the perception of falling asleep along with other perceptual and cognitive variables, such as altered reality-remoteness, low familiarity, sudden ideas without goal-orientation, and lack of body perception (Lehmann, Grass, & Meier, 1995). Another study using principle component analysis identified a dimension of subjective experience that differentiated physiologically defined sleep states (stage 1 and stage 2) from waking states. The experiences that had the highest loading included the loss of awareness of the experimental situation, reported sleepiness, and inward directed attention (Wackermann, Pütz, Büchi, Strauch, & Lehmann, 2002). Although these studies did examine subjective experience during sleep onset, their main concern was to compare dream mentation to sleep onset experience, or to search for associations between subjective experiences and EEG activity. They did not attempt to identify the subjective experiences that determine the perception of falling asleep.

Previous studies have consistently demonstrated that sleep onset processes are associated with a decrease in awareness of environmental stimuli and with diminishing thought control. But they have not clearly identified precisely what factors are involved with the perception that we have fallen asleep. The primary goal

of the current study is to probe the subjective experiences that are critical to this perception. Subjects were awakened at several junctures during sleep onset, in order to identify the various subjective experiences involved. Regression analyzes were then conducted in order to tease out those experiences that are decisive in explaining perception of sleep onset.

2. Methods 2.1. Subjects

Twenty-six subjects were recruited from a university campus, to participate in the study. Because six of them could not fall asleep after repeated awakenings, they were excluded from data analysis. Twenty subjects (nine males and 11 females), completed the procedures. Their mean age was 24.2 years, with a standard deviation of 3.24. The criteria for inclusion were: (1) age between 20 and 35 years; (2) no current or past major medical or psychiatric illnesses, and no evidence of sleep disorders; (3) no current use of prescribed or leisure drugs that might affect sleep; and (4) non-shift workers with regular sleep– wake schedules. Potential subjects were screened for sleep, psychiatric, and major medical disorders in a clinical interview conducted by a trained, clinical psychology, graduate student.

2.2. Procedure

The subjects who satisfied the inclusion criteria were scheduled to arrive at the sleep laboratory, based in a university, for an afternoon-nap test. In order to enhance the likelihood that subjects would continue to fall asleep, even after repeated experimental awakenings during the nap, they were instructed to sleep 2 h less than usual during the night prior to their arrival at the laboratory. After obtaining the subjects' informed consent, electrodes were attached. While the electrodes were being positioned, the subjects were instructed to read through a list of questions that would be asked immediately after each awakening during the course of the experiment, questions used to assess their subjective experiences. The meaning of these questions was made clear to the subjects prior to the start of the experiment. Before beginning the nap test, the subjects were informed that an intercom system would be used to awaken them, and that their interview would be taken immediately afterward.

The recording montage for the polysomnography (PSG) included: an electroencephalogram (EEG), with electrodes placed at C3/A2, C4/A1, O1/A2, O2/A1; electrooculography (EOG) to measure left and right eye movements; chin electromyography (EMG); and electrocardiography (ECG). The impedances of all electrodes were kept below 5 kX prior to the start of recording. A total of five awakenings were conducted for each subject. The first awakening was a practice run,

to familiarize the subjects with the procedures. It was conducted 2 min after the light was put out, regardless of whether the subject was in the wake state or sleep state. Data from the first awakening were not analyzed. The other four awakenings were conducted on the basis of different PSG features, which were as follows:

The emergence of a clear slow eye movement (SEM). The onset of stage 1 sleep, defined as the first 30-s epoch in which EEG alpha activities decreased to less than 50% (S1). The onset of stage 2 sleep, as defined by the emergence of a K-complex or a sleep spindle (S2) A 5-min continuation of stage 2 sleep (S2+5).

In order to avoid sequence effects, sequencing of the four awakening junctures was counter-balanced across all subjects. The awakening junctures were identified on-line by a well-trained graduate student and were independently confirmed by another well-trained graduate student.

Each time the subject was awakened, his or her name was called out through an intercom system. As soon as the ongoing EEG display indicated that the subject was fully awake, an experimenter entered the bedroom and conducted the interview while the subject lay in bed. During the interview the lights remained off, such that the room was just faintly illuminated by light from the adjacent monitoring room. The experimenter assessed the subject's perception of the sleep state through a structured questionnaire designed to probe various aspects of subjective experience. Questions were as follows:

1. Perception of Sleep

- 1-1. Sleep Perception: "Did you fall asleep?" (Y/N)

- 1-2. Depth of Sleep: "How deep was your sleep?" (0-5)

2. Experiences of Sensation and Perception

- 2-1. Clarity of Environmental Perception: "How clearly were you able to perceive any environmental stimuli?" (0-5)
 - 2-2. Visual Image: "Did you see any visual images?" (Y/N)

- 2-3. Vividness of Visual Image: "How vivid was the visual image?" (0-5)

- 2-4. Auditory Perception: "Did you hear any sounds and/or voices?" (Y/N)

- 2-5. Clarity of Auditory Perception: "How clear were the sounds/voices?" (0-5)

- 2-6. Other Sensory Experiences: "Were there any other sensations such as bodily or olfactory sensations?" (Y/N)

- 2-7. Control over Perception: "Were you able to control your perceptual experiences?" (0-5)

4-1. Perceived Reality: “How real did any of the experiences seem to you?” (0–5) (This question is classified under Ori-entation and Involvement, but was located here on the questionnaire.)

3. Thinking Processes

3-1. Thinking Experience: “Were you thinking of anything when I called out your name?” (Y/N)

3-2. Control over Thinking Process: “How well were you able to control your thoughts?” (0–5)

3-3. Coherence of Thinking Process: “How coherent was your thinking process?” (0–5)

3-4. Logic of Thinking Process: “Were your thoughts logical?” (0–5)

4-2. Daily-life concerns: “Were the thoughts related to your daily-life concerns?” (0–5) (This question is classified under Orientation and Involvement, but was located here on the questionnaire.)

4-3. Here-and-now experience: “Were the thoughts related to the situation in the lab?” (0–5) (This question is classified under Orientation and Involvement, but was located here on the questionnaire.)

4. Orientation and Involvement

4-4. Sense of Involvement: “Did you feel more like an observer (0), or did the thoughts and experiences seem to be yours (5)?” (0–5)

4-5. Orientation: “To what degree were you aware that you were in the lab and lying in bed?” (0–5)

5. Emotion

5-1. Emotional Experience: “Were you experiencing any emotion at the moment I called you?” (Y/N)

5-2. Types of Emotion: “What type of emotion did you experience?”

5-3. Valence of Emotion: “Was the emotion positive or negative?”

5-4. Intensity of Emotion: “How intense was the emotion?” (0–5)

3. Data analysis

Subject responses on the yes/no questions were analyzed first. Chi-square was used to analyze answer frequencies for the four junctures of awakening. Ratings for intensity of the four conditions were then compared. Since the ratings were on Likert-type scales, which are ordinal, non-parametric statistics were used. A Friedman test was employed to compare the ratings among the four conditions. And, a

Wilcoxon signed rank test along with Bonferroni’s correction, were used to make post hoc comparisons.

Bivariate Spearman’s rank correlation analyzes were first conducted to identify the variables that are associated with subjective estimates of sleep depth. In order to identify the experiences that determine the subjective perception of having fallen asleep, a forward stepwise logistic multiple regression was then conducted to identify the predictors for the response to the binary question “did you fall asleep?”

4. Results

As expected, the average durations of time for the subjects to reach the four junctures of awakening increased from SEM through to S2+5: they were, respectively, 189.0 s, 411.3 s, 605.9 s, and 1162.0 s. The frequencies of various subjective experiences are presented in Table 1. Chi-square analyzes show that among the several conditions only Sleep Perception differed. As expected, the perception of having fallen asleep increased throughout the sleep onset period. The Friedman’s test – comparing the ratings on different dimensions of subjective experience across the junctures of sleep onset – reveals that most of the ratings (sleep depth, clarity of environmental perception, control over perception, control over thoughts, coherence of

Table 1
Frequencies, percentages, and Chi-square results for the yes/no questions.

Item	Juncture of Awakening					X ²	df	p
	Y/N	SEM	S1	S2	S2+5			
Sleep perception	Y	3(15%)	8(40%)	9(45%)	16(80%)	17.37	3	.001
	N	7(85%)	12(60%)	11(55%)	4(20%)			
Visual image	Y	6(30%)	13(65%)	11(55%)	11(55%)	5.35	3	.148
	N	14(70%)	7(35%)	9(45%)	9(45%)			
Auditory perception	Y	10(50%)	7(35%)	4(20%)	4(20%)	5.76	3	.124
	N	10(50%)	13(65%)	16(80%)	16(80%)			
Other sensory exp.	Y	10(50%)	6(30%)	4(20%)	4(20%)	5.71	3	.126
	N	10(50%)	14(70%)	16(80%)	16(80%)			
Thinking experience	Y	14(70%)	11(55%)	9(45%)	9(45%)	3.37	3	.338
	N	6(30%)	9(45%)	11(55%)	11(55%)			
Emotional experience	Y	5(25%)	2(10%)	5(25%)	0(0%)	7.06	3	.070
	N	15(75%)	18(90%)	15(75%)	20(100%)			

thoughts, logic of thoughts, perceived reality, sense of involvement, and orientation) changed progressively, in step with the appearances of physiological indices of sleep, from SEM through to S2+5. Significant differences, however, occur at different junctures, for different aspects of subjective experience. It appears that most changes in sensory experience occur early in the process; later, only slight changes occurred. Changes in the thinking process also started early, but continued to change significantly even during the latter stages of sleep onset. Orientation and involvement, on the other hand, did not exhibit significant change until after the start of S1, changes which continued throughout the entire process (see Table 2 and Fig. 1).

Finally, a few aspects of subjective experience did not evince any significant differences at the different junctures. Specific sensory experiences, such as the vividness of visual imagery, the clarity of auditory perception, and emotional intensity, did not significantly differ among the four conditions.

Table 3 presents the Spearman's correlation coefficients among the variables and the subjective estimates of sleep depth. As indicated by the table, significant correlations were found for most of the subjective experience variables, except for vividness of visual imagery, daily-life concerns, and intensity of emotion. All the variables that showed significant correlations were regressed onto Sleep Perception in a stepwise fashion. Logistic regression identified control over thinking process and logic of thinking as the variables that predict sleep perception. These two variables could explain about 66% of variance in the perception of sleep ($X^2(2) = 54.16$, $p < .001$; Nagelkerke R square = .658). Control over thought processes and logic of thinking could correctly classify 91.7% of the perception of sleep and 84.1% of the perception of waking. The overall correct rate was 87.5% (see Table 4).

5. Discussion

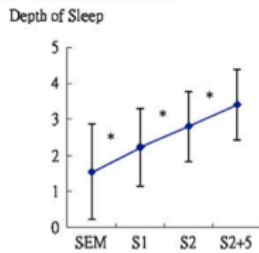
This study aims to explore the subjective experience of the sleep onset process and to identify those experiences that are specifically associated with the perception of having fallen asleep. As expected, the perception of falling asleep as well as subjective estimates of sleep depth increased at each of the four junctures of awakening. Just as has been suggested by previous studies, consistent reports of having fallen asleep were not obtained until 5 min after immersion into stage 2 sleep. Most aspects of subjective experience also changed progressively during the course of sleep onset, but at different paces for different domains. The perception of environmental stimuli dropped significantly from SEM through stage 1 sleep. But after the onset of stage 2, there was little further decline. Thought process was also shown to degenerate progressively during the course of sleep onset. Levels of control, coherence, and logic of thought declined, from the start of SEM, but significant changes did not appear until the start of stage 2 sleep; decline continued even more after the subject began to sleep soundly. Orientation and perceived reality, on the other hand, showed no significant change from SEM to stage 1 sleep, but did change significantly after the inception of stage 2. In other words, significant change did not appear until the latter stages of sleep onset. Emotional intensity was low from the start and evinced little difference at any of the awakening junctures. These

Table 2
Means, standard deviations (SDs), and the results of Friedman's test comparing subjective experiences among different junctures of awakening, and post hoc tests with Wilcoxon signed rank test.

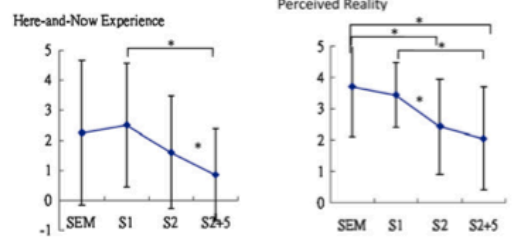
Variable	SEM(A)		S1(B)		S2(C)		S2+5(D)		X ²	Post hoc
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
<i>Perception of Sleep</i>										
Depth of Sleep	1.55	1.32	2.23	1.07	2.80	.97	3.40	.98	33.06***	A < B < C < D
<i>Sensation and perception</i>										
Clarity of Environmental Perception	3.35	1.50	2.00	1.42	1.68	1.47	1.05	1.00	28.50***	A > B = C = D
Vividness of Visual Image	.90	1.68	1.80	1.64	1.15	1.46	1.15	1.39	4.61	
Clarity of Auditory Perception	1.09	1.79	.75	1.48	.53	1.19	.20	.52	4.63	
Other Sensory Experiences	1.43	1.93	1.20	1.94	.60	1.50	.60	1.27	4.18	
Control over Perception	2.90	1.86	2.33	1.59	1.70	1.45	.83	1.39	17.16**	A = B = C > D
<i>Thinking processes</i>										
Control over Thinking Process	3.40	1.70	2.25	1.74	1.33	1.26	.63	.81	30.85***	A = B, A > C > D, B = C > D
Coherence of Thinking Process	2.75	2.07	1.53	1.76	.78	1.32	.40	.88	17.57**	A = B > C, A > D, B = D, C = D
Logic of Thinking Process	3.35	2.03	1.90	2.10	1.30	1.75	.60	1.05	20.98***	A = B = C, A > D, B > D, C = D
<i>Orientation and involvement</i>										
Perceived Reality	3.70	1.59	3.43	1.04	2.43	1.52	2.05	1.64	21.57***	A = B > C = D
Daily-Life Concerns	2.20	2.19	2.90	1.68	2.45	2.09	1.65	1.84	8.05*	A = B = C, A = D, B > D, C = D
Here-and-Now Experiences	2.25	2.40	2.50	2.06	1.60	1.88	.85	1.53	12.59**	A = B = C, A = D, B > D, C = D
Sense of Involvement	3.95	1.54	3.10	1.97	2.48	2.07	1.68	1.76	15.07**	A = B, A > C = D, B = C = D
Orientation	4.05	1.70	3.33	1.78	2.33	1.89	1.25	1.55	34.52***	A = B > C > D
<i>Emotion</i>										
Intensity of Emotion	.73	1.33	.48	1.27	.45	1.05	0	0	7.15	

* $p < .05$.
 ** $p < .01$.
 *** $p < .001$.

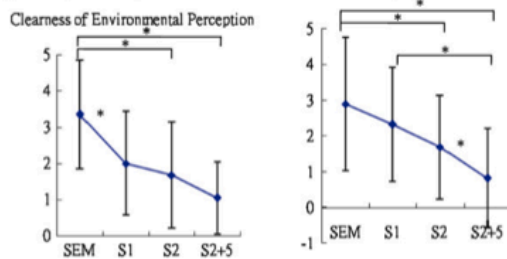
(A) Perception of Sleep



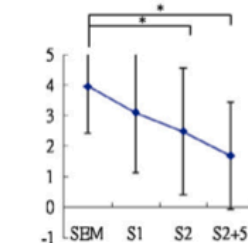
(D) Orientation & Involvement



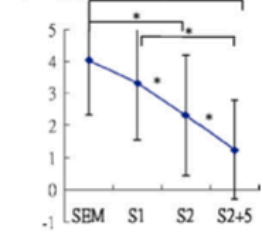
(B) Perceptual Experience



Sense of Involvement



Orientation



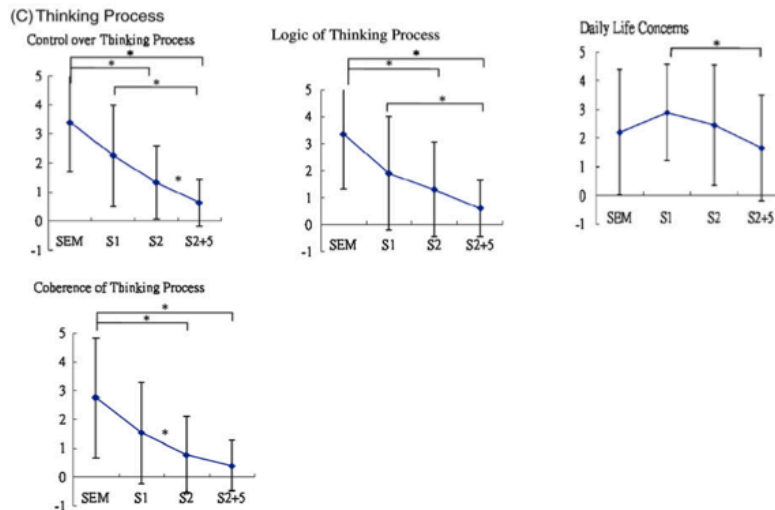


Fig. 1. The ratings on subjective experiences with statistically significant changes over the four junctures (SEM: slow eye movement, S1: stage 1 sleep, S2: stage 2 sleep, and S2+5: 5 min of stable stage 2 sleep) during sleep onset period. Asterisk represents significant difference on Wilcoxon signed rank test with Bonferroni's correction.

findings are consistent with previous studies that revealed significant change in sensory and thought process during sleep onset, while emotional experiences maintained a low intensity throughout (e.g., Foulkes & Vogel, 1965; Gibson et al., 1982; Hori, Hayashi, & Hibino, 1992; Hori, Hayashi, & Morikawa, 1991).

This study's most significant finding is that control over the thinking process is the most prominent subjective experience that is associated with the judgment of having fallen asleep. Thought logic was also identified as a predictor for falling asleep. In addition, our study showed that although sensory processing decrease and thinking process degeneration began early in sleep onset, degeneration continued after immersion into light sleep and more stable sleep. But perception of environmental stimuli showed relatively less diminution. And, degree of orientation was not significantly affected until after stage 2 sleep had been achieved.

Our results may help explain the findings of previous studies which reveal that subjects did not perceive sleep onset until after stable, stage 2 sleep had been achieved. As mentioned above, reports of having slept were more closely associated with

Table 3
Spearman correlation coefficients among different domains of subjective experience with subjective estimates of depth of sleep.

	Depth of sleep	p
<i>Experiences of sensation and perception</i>		
Clarity of Environmental Perception	-.596***	<.001
Vividness of Visual Image	.170	.131
Clarity of auditory perception	-.235*	.036
Other Sensory Experiences	-.208	.064
Control over Perception	-.583***	<.001
<i>Thinking processes</i>		
Control over Thinking Process	-.685***	<.001
Coherence of Thinking Process	-.527***	<.001
Logic of Thinking Process	-.605***	<.001
<i>Orientation and Involvement</i>		
Perceived Reality	-.502***	<.001
Daily-Life Concerns	-.184	.102
Here-and-Now Experiences	-.277*	.013
Sense of Involvement	-.305**	.006
Orientation	-.625***	<.001
<i>Emotion</i>		
Intensity of Emotion	-.251*	.025

*** Correlation is significant at the .001 level (2-tailed).

** Correlation is significant at the .01 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table 4
The percentage of correct prediction of perceived sleep with the predictors identified using logistic multiple regression.

	Observed	Predicted		Percentage correct
		Awake	Asleep	
Model 1	Awake	37	7	84.1
	Asleep	5	31	86.1
	Overall percentage			85.0
Model 2	Awake	37	7	84.1
	Asleep	3	33	91.7
	Overall percentage			87.5

the occurrence of stage 2 sleep than with stage 1 sleep. Perception of sleep was previously reported to be related to substantially diminished responsivity to visual and auditory stimuli (Agnew & Webb, 1972). But Bonnet (1986) demonstrated that although the auditory threshold increased soon after the appearance of the sleep spindle, perceptions of having fallen asleep were not reported until several minutes later. Bonnet's findings are in line with our results, that the perception of sleep onset is more associated with deteriorating thought processes than with perception of external stimuli. One might perceive oneself as being awake, even after the perception of environmental stimuli abates. What seems to be required for the perception of having fallen asleep is substantial loss of control over the thought process along with deterioration of logical reasoning, phenomena that obtain after one enters a period of sustained stage 2 sleep.

As Rechtschaffen (1994) stated, "there are separate effector mechanisms that control the different behavioral conditions that, when considered together, constitute 'sleep'." Our results also support the finding that sleep onset is not a single process; rather, it is parallel processes that comprise multiple components. Changes in different aspects of subjective experience may reflect different underlying mechanisms. The different mechanisms may operate at different rates, thereby corresponding to differential rates of change in subjective experience. Some of the

mechanisms may be reflected in the physiological phenomenon that we use to define sleep onset or sleep stages, while others may not. Previous studies have shown that the decline in EEG alpha activity is associated with the subjective report of loss of awareness of the environment (Davis et al., 1937); this is consistent with our findings that the onset of stage 1 sleep is associated with a decline in perception of environmental stimuli. A recent study also showed that the clarity of mental content and controllability of thought both diminish after the onset of stage 2 sleep (Weigand, Michael, & Schulz, 2007). In our study, control over thinking was found to continue to decline throughout stage 1 and stage 2 sleep, and orientation and involvement were found to decrease after the inception of stage 2 sleep.

Recent studies have also combined EEG and brain imaging techniques in order to explore changes in region-specific brain activity during sleep. The findings obtained from these studies suggest a global decrease in cerebral and thalamic activity during non-rapid eye movement (NREM) sleep. But the specific brain regions identified as responsible for the changes varied somewhat. Generally speaking though, light sleep was associated with decreased activity in the frontal and parietal areas of the cortex and in the thalamus. Deep sleep is characterized by a further decrease in activity in these areas, as well as within the basal ganglia (e.g., Balkin et al., 2002; Braun et al., 1997, 1998; Finelli, Baumann, Borbély, & Achermann, 2000; Kjaer, Nowak, & Lou, 2002; Maquet, 2000; Nofzinger, Mintun, Wiseman, Kupfer, & Moore, 1997; Nofzinger et al., 2002; Peigneux et al., 2001). As for stage 1 sleep, Czisch and his colleagues reported fMRI indications of reduced activation in both the auditory and visual cortex in response to auditory stimuli, as demonstrated by a decrease in the blood-oxygenation (Czisch et al., 2002). Also concerning stage 1 sleep, Kaufmann and colleagues' (2006) fMRI studies indicated decreased activation in the thalamic and cingulate structures, other limbic areas, frontal lobes, occipital lobes, temporal lobes, and the insula. These changes may be responsible for a diminution of sensory and thought processes after achieving stage 1 sleep. Stage 2 sleep, on the other hand, was associated with decreased activity in the thalamic and hypothalamic regions, cingulate cortex, right insula and adjacent regions of the temporal lobe, the inferior parietal lobule, as well as the inferior and middle frontal gyri (Kaufmann et al., 2006). These studies when coupled with our findings suggest that diminution of activity in these areas may be responsible for the loss of orientation and self awareness that occurs in stage 2 sleep. Unfortunately though, no detailed subjective experiences were

assessed in these brain imaging studies. Knowledge of the relevant neural correlates of these subjective experiences must await future studies.

It should perhaps be noted that the changes in degree of control over and in the logic of thought documented here are not necessarily sustained throughout the entirety of sleep. Several recent comparisons of waking and dream thought have motivated a refinement of prior distinctions: neither, it seems, is dream cognition as inherently deficient in the ways that some have argued, nor is waking cognition as proficient. According to a moderate revision of the distinction, dream thought should not be understood monolithically (Kahn & Hobson, 2005). For example, there appear to be two distinct types: one employs context logic, which reasons from premises; the other, state logic (metacognition), which reasons about premises. It is only the latter that is absent in the dream state. According to a more robust revision of the distinction, dreaming and waking cognition do not differ qualitatively (Kahan, LaBerge, Levitan, & Zimbardo, 1997). Reflective awareness and other metacognitive experiences might well be more common in dream cognition than is typically believed; that such experiences are seldom reported might be more a reflection of methodological artifact than of actual dream experience. Possible differences between moderate and more robust revisions of the distinction between waking and dream thought need not be adjudicated here. It need only be pointed out that both sets of studies suggest that some measure of what is lost when subjects perceive the onset of sleep is later recovered when dream cognition commences.

A possible limitation of the current study is that it was conducted on daytime naps rather than nocturnal sleep. It is perhaps arguable that correlations between subjective experiences of sleep onset and physiological indices of sleep onset differ, depending upon whether one is napping or engaged in nighttime sleep. Although such a view is neither motivated by current theory nor by empirical evidence, to rule out this possibility further, supplementary studies should be carried out.

G. Results and Discussion, Part III: Belief and Ethics

i. The ethics of false belief.

I. Introduction

Allen Wood (2002, 2008) argues that the main principle governing the ethics of belief is the “procedural principle.”¹ According to this principle we should apportion the strength of our beliefs to the evidence. In other words we should believe “only what is justified by the evidence, and believe it to the full extent, but only to the extent, that it is justified by the evidence” (2008: 9). Wood qualifies this claim in certain respects (2008: 13, fn. 8) and, grudgingly, acknowledges the possibility of non-trivial exceptions (2002: 38). Nevertheless, he concludes that failing to adhere to this principle invariably violates our self-respect and is, as well, in other regards inevitably corrupting (2002: 36, 2008: 24). Similar sentiments have been expressed by other philosophers, such as Michael Lynch (2004: 143), who writes: “Caring about truth and believing the truth about what you care about are necessary parts of happiness by being necessary parts of integrity, authenticity, and self-respect.”

Wood is aware of the body of empirical work which suggests that people benefit from holding certain false beliefs, as well as beliefs not supported by evidence. But for various reasons he denies that this work counts against the procedural principle. For example, he (2008: 13, fn. 8) proclaims that “no one could stably hold both the belief that is supposed to benefit them and also know that it is false . . . even if illusions do benefit people’s health, it does not seem that this is justification a person could stably or self-consistently apply to their own beliefs.” Note that Wood seems to be making an empirical claim about the nature of beliefs.

Wood’s views as regards both normative and descriptive aspects of belief are consistent with the received view of beliefs, viz. that they “aim at the truth” (Williams, 1973: 137-138).² Many, perhaps a majority, of late 20th and early 21st century philosophers have converged on the view that beliefs are constituted in such a way that they can be accurately characterized by this phrase. Donald Davidson emphasizes their “veridical nature” (2003: 366-367) and he (1977: 295) argues that “successful communication proves the existence of shared and largely true, view of the world;” John R. Searle (2001: 37-38, 257) claims that it is their “job” to “represent how things are;” Peter A. Railton (2003: 297) holds that belief “not only represents its propositional content as true,” it “cannot represent itself as unresponsive to—unaccountable to—their truth;” Tim Crane (2001: 103) says that “holding true” is a synonym for belief; Bernard Williams (2002: 80) claims that beliefs are “subject to a norm of truth;” and, Ralph Wedgwood (2002: 273) observes that “for every proposition *p* that one consciously considers, the best outcome is to believe *p* when *p*

is true.” Wood’s view, along with this cluster of interrelated views, I refer to as the Truth-Tracking View (TTV) of belief.

I shall not concern myself with strong versions of TTV, for their vulnerabilities are conspicuous. I here consider only modest versions, of which I take J. David Velleman’s (Shah & Velleman, 2005; Velleman, 1999, 2000) to be representative. Both Wood and Velleman exemplify TTV, but Velleman provides a more detailed account. Moreover, he is in sympathy with Wood’s normative position, and is sensitive to relevant criticisms of TTV.

Velleman’s version is used below, in part, as a foil against which to develop the idea of “aiming away.” Beliefs that aim away from the truth I refer to as “Tertullian beliefs” (“t-beliefs”). Although all modest versions of TTV do qualify the sense in which beliefs can be said to aim at the truth (hence, the attributive “modest”), still they lack the resources with which to account for the distinctive causal-explanatory role played by t-beliefs. The standard qualifications that are appended to modest versions of TTV, while perhaps succeeding in making them weakly compatible with instances of t-belief, also make it appear that t-beliefs are nothing but incidental, variously inconsequential, or “pernicious” (Wood, 2002: 40), features of our cognitive economy. After sketching Velleman’s account, t-belief is introduced by means of examining certain commonplace, anecdotal instances wherein behavior contravenes professed beliefs in ways that suggest belief-forming mechanisms are not responsive to evidence in the ways required by TTV. Next empirical studies of beliefs that aim away from the truth are reviewed. Then the distinctive characteristics of t-beliefs are limned, such that they can be clearly distinguished from other attitudes like desire, hope, or hypothesis. Finally, I evaluate Wood’s procedural principle in light of what we are now learning about t-beliefs. I argue, pace Wood, that the capacity to occasionally and strategically aim askew of the truth might be essential to the maintenance of self-respect and that it is not necessarily corrupting in the ways that he suggests. In a brief concluding section I suggest that whether or not Wood is correct in his uncompromising advocacy of the procedural principle will ultimately be determined by the results of empirical research—sociological, psychological, and neuroscientific.

II. A Modest Version of TTV

On Velleman’s (2000: 255) version of TTV, belief is constituted both “by its power to cause behavioral output,” and by “its responsiveness to epistemic input.” It is not sufficient to claim that belief takes its propositional object as representing the

way things are, for this alone could not distinguish it from certain other attitudes (Velleman, 1999: 198-200). What distinguishes belief from, say, assumption or imagination is “the spirit” in which a propositional object is regarded as true: an assumption might be “tentatively” held and something imagined might be “fancifully” held, but a belief is “seriously” held. Fantasies and assumptions are not “serious” because they entail accepting a proposition as true without sensitivity to whether a person is “accepting” the truth. To believe is not to “accept” for polemical or heuristic purposes (as is the case with assuming), neither is it to “accept” for recreational or motivational purposes (as is the case with imagining); instead, to believe is to accept a proposition “with the aim of doing so if and only if it really is true” (Velleman, 1999: 200; see also Wood, 2002: 19-20).³ Beliefs are regulated—formed, revised, and extinguished—in truth-conducive ways, in ways that are responsive to evidence and reasoning (Shah & Velleman, 2005: 498).⁴

To say that beliefs aim at the truth is not to say that the aim is to believe as many truths as possible; nor is the aim to believe as many as possible useful or valued truths; nor indeed is it to say that the aim is maximizing the proportion of truths to falsehoods among one’s beliefs (Velleman, 2000: 251-255). TTV requires only that beliefs aim at the truth in some way, while allowing that there are multiple ways in which they might do so. It further allows that belief is not exclusively governed by truth-seeking mechanisms (Shah & Velleman, 2005: 500-501; Velleman, 2000: 254): some mechanisms may cause beliefs that occasionally diverge from the truth (the adoption of better-safe-than-sorry strategies). But Velleman holds that belief is necessarily subject to mechanisms designed to make it true: “the input constraints definitive of belief are designed to yield beliefs that are true” (Velleman, 2000: 277). In other words, mechanisms that are not truth-seeking are not definitive of belief.

Because only some beliefs are caused by the goal-directed activity of persons, and many are the results of processes that do not involve agential goals or intentions, the concept of belief must include more than just the manner in which beliefs are actually regulated. A “standard of correctness,” a normative standard, must also be applied (Shah & Velleman, 2005: 498-500). Modest TTV then conjoins the descriptive and the normative: belief is regarded as “truth-regulated acceptance.” And to this a norm of truth is then applied. Norms governing belief are understood in a “biconditional” sense (Shah & Velleman, 2005: 519): although they do not require acceptance of every belief that would be correct, they do forbid the holding of a belief that would be incorrect. These norms may be lax in what they require a person to

accept, but they are strict in what they prohibit—the holding of incorrect beliefs. Wood (2008: 10) expresses this prohibition thus: “beliefs not justified by the evidence are immoral.”

Velleman (2000: 277-279) considers the possibility that beliefs might aim at “instrumental success” or “empirical adequacy.” But he claims that while we might sometimes settle for an alternative to truth as a “second-order” aim, truth remains the “first-order” aim. Suggesting an analogy he observes that a basketball player might proclaim that his ultimate aim is to earn a salary increase, but fans don’t thereby presume that everything he does on the court is aimed at the salary increase, because the best way to achieve the salary increase is to aim at victory itself. Money might be the object for playing the game, but within the context of the game winning is adopted as the aim. “Similarly, we may enter the game of having beliefs on a particular subject because we want our motivating cognitions on that subject to yield successful actions; but success in action does not thereby become the object of the game.”

Velleman (2000: 278) is dismissive of the possibility that we might discover beliefs to be regulated so as to aim at something other than truth. He justifies this dismissal with a claim about the content of our introspections: when we discern a gap between a belief and the truth, the belief becomes unsettled and starts to change (see also Wood, 2008: 13, fn. 8). Alternatively, if the belief persists, another belief is formed to help close the gap, while the original belief is reclassified as an illusion or bias. Non-evidential considerations simply cannot be explicitly treated “as relevant to the question what to believe. Any influence that such considerations exert must be unacknowledged.”⁵

III. Some Doubts About TTV

If, as Velleman contends, beliefs are constituted by their power to cause behavior and by their responsiveness to epistemic input, if they aim at the truth, and if they are indeed subject to mechanisms or constraints that are designed to yield true beliefs, then, at minimum, we might reasonably expect that when, on good evidence, a person categorically and sincerely asserts that a belief is untrue, that same person should not act as though it were true, especially when acting as though it were true incurs greater cost than would be incurred were one to act in accord with professed beliefs. But there seem to be clear cases in which people possess the relevant evidence, do so assert, and yet act as though they hold beliefs which they deny holding, even when doing so carries significant cost. And these are not, and are not relevantly similar to, cases of assumption or fantasy. What’s more, they at least seem to

allow for the possibility that non-evidential considerations are being explicitly treated as relevant to the question what to believe. At least it is not obvious that these considerations exert their influence in a manner that is wholly unacknowledged.

Consider the case of a person born and raised in the western world who categorically denies believing that the number 13 invites bad luck.⁶ Moreover, this very same person is familiar with the arguments and the vast amount of evidence that demonstrate that 13 is no more lucky or unlucky than any other number. Nevertheless, given the option of choosing between a hotel room or an office space on the 13th or the 14th floor, all things being equal, he might well be disposed to choose the latter. What's more, for many people the same would likely hold true even if all things weren't equal; that is, even if avoiding 13 were to require greater cost. Doubtless there is a limit to just how much greater cost one would be willing to incur in order to avoid the 13th, but the expenditure of significant time, money, and other resources in the avoidance of 13 is not uncommon.

Were we to employ belief-desire psychology toward explaining the relevant behavior, we would likely say "Stan's belief that 13 is unlucky caused him to choose the 14th rather than the 13th floor," as part of our explanatory sketch. Here then we would have a reasonably clear case in which one has, for good reason, denied believing that 13 is unlucky, yet, a specific decision was prompted, *inter alia*, by just that very belief.⁷ While the relevant psychology is insufficiently understood, it is by no means obvious that non-evidential considerations exert influence in a manner that is wholly unacknowledged.

Recall that, according to Wood and Velleman, if we discern a gap between a belief and the truth, the belief becomes unsettled and starts to change, or another belief is formed to help close the gap, while the original belief is reclassified as an illusion or bias. But in this case it seems plausible to claim that the person can be aware of the gap and that the belief neither becomes unsettled, nor does it require formation of a gap-closing belief. A balanced perspective should allow that post hoc introspection concerning cases of this sort is theory-laden reflection over "skittish" phenomena (Hurlburt&Schwitzgebel, 2007: 48-53). Hence, a charitable view of the Wood-Velleman position is that we are left with an introspective stand-off.

If cases of superstition strike some readers as too exotic, perhaps vanity products can serve as more compelling examples. Alleged cures and treatments for alopecia (and the many other assaults on personal vanity) are as numerous as their evidence is wanting. But just as with superstitions, there are many people who categorically deny

believing that these products can promote good health or restore one to a hirsute state, yet they act as though they hold the very beliefs they deny holding, even when doing so carries significant cost.⁸ Not only are the claims unsubstantiated, positive reasons not to believe the claims are plentiful; yet, intelligent consumers behave in ways that contravene professed beliefs. As is the case with superstition, here too the possibility of acknowledged, non-evidential considerations playing a role in belief regulation cannot be dismissed out of hand.

Some contexts, in particular those that are harrowing or life-threatening, can help to further illustrate the point about superstition and about marketing gullibility. Consider the case of a medical doctor (or scholar, or scientist) who is quite convinced on extremely good evidence that herbal treatments like echinacea cannot prevent the common cold and prayer cannot cure bone cancer.⁹ But when the throat begins to feel raw, or while awaiting the results of the biopsy, some among these very same people are highly disposed to purchase echinacea or stop by a temple, church or synagogue. As for herbs like echinacea, since the common cold is typically just a nuisance, one might wonder why the person who categorically denies believing in its effectiveness would be so easily motivated to behave in accord with the belief that it is effective.¹⁰ The threat of bone cancer is of course another matter though: desperate to cling to life one might grasp at any measures, no matter how far-fetched, and without regard to whether the person has spent a lifetime emphatically not believing in the method that is being tried. Desperation trumps justification.

Perhaps it might be argued that when confronted by life-threatening illness we abruptly adopt an assumption, an attitude that need only be “tentatively” held, like a heuristic, in order to motivate experimentation with herbs or prayer. But to characterize this attitude as a heuristic, after a life-time of deliberate, well-considered, disbelief, would be odd. Typically assumptions are adopted as a means of exploring the unfamiliar, in an attempt to gain new knowledge. But that does not seem to be a straightforward characterization of what is happening here, for by hypothesis these are cases concerning which the person previously explored the relevant claims and, for good reason, rejected them. It seems more natural to say that desperation alters belief regulation such that one becomes strongly influenced by non-evidential considerations. If this latter characterization is correct, then one explicitly treats non-evidential considerations as relevant to the question what to believe.

Velleman allows that people will sometimes choose to error on the side of caution, as when worried about potential predators. But this does not help us to

explain the avoidance of 13 or the abrupt decision to behave in accord with beliefs that one rejects. Many people are familiar with the relevant evidence (we might suppose them to be avid readers of the *Skeptical Inquirer* and like material), so, unlike wilderness predation, there simply is no reason to be wary of 13 and no reason to suddenly embrace prayer. Be that as it may, people do behave in these ways.

IV. The Empirical Study of False Beliefs

Proper characterization of the preceding examples is contentious. They are anecdotal and their interpretations, uncertain, in part due to the vagaries of introspection. But the limitations of introspection do not imply that an interpretive stalemate is inevitable. There are some well-studied examples of belief that systematically diverge from the truth in ways which put pressure on the TTV characterization; especially worthy of note are the “positive illusions.” These have been variously described and classified but, according to one of the better known sets of studies, they include self-aggrandizing perceptions, illusions of control, and unrealistic optimism (Taylor & Brown, 1988, 1994).¹¹

Shelley E. Taylor and Johathon D. Brown have amassed considerable evidence to suggest that people consistently see themselves in a more positive light; others, in a negative light, relative to self. In commenting on this, the “better-than-most” effect, they observe that it is difficult if not impossible for any one to be warranted in believing that he is, for example, kinder, warmer, more humorous and more sincere than the average person. As regards illusions of control, the claim is not that people believe themselves capable of exercising control over that which clearly exceeds their reach; rather, this is a moderate distortion concerning those things over which people are in fact able to exert some control. And, there is a voluminous body of literature testifying to the claim that most people are unrealistically optimistic in believing their future will be better than can be justified on statistical grounds.

In effect, people tend to believe in a self-image that reassures. People consistently overestimate their abilities, whether in matters of leadership, getting along with others, or even just driving skills. These tendencies are not merely widespread among the poorly educated; as many as 94% of university professors assessed themselves as better at their jobs than their “average” colleagues (Cross, 1977). Moreover, most people, even when provided with accurate, relevant base rate information, tend to underestimate the likelihood that they will be stricken with cancer, be in a car accident, get divorced, and so forth. Pronin (2008, 2007; Pronin&Kugler, 2007; Pronin, Lin, & Ross, 2007) has devoted special attention to

this final point. She discovered that when subjects are informed about “introspective illusions” and “bias blind spots” they, nevertheless, adjudge themselves to be less susceptible than others. Even when subjects—immediately after acting in accord with a particular bias—are presented with an explicit description of the bias, a description that indicates it is a common human tendency, they still fail to see themselves as liable. And these results are not indications of reticence, for instructions given to subjects make it clear that experimenters want to know whether bias is present and make it clear that the bias is common (Pronin, Lin, & Ross, 2002: 375).

Emily Pronin and Matthew B. Kugler (2007) have found that the only way to prompt subjects to recognize personal vulnerability is to specifically educate them concerning the epistemic failings of introspection. It remains unknown though whether such focused education can bring about efforts to compensate for bias (Pronin, 2007: 40). At the very least compensation would be difficult: recent evidence shows that the way we think about self in the present differs substantially from the ways in which we think about past or future selves. The limbic system and, consequently, affect, is much more engaged when people think about themselves in the present (Pronin, 2008:1179-1180). This suggests that anticipations of the future or post hoc interpretations might be correctable in ways that judgments about the self-at-this-moment are not.

Does this evidence unequivocally demonstrate that Velleman is wrong in claiming that only unacknowledged non-evidential considerations can affect belief regulation? Not necessarily. Since the strategy here is not to cherry-pick results, it must be admitted that some evidence suggests, at least for individual events, after carefully being instructed concerning the frailty of introspectively based knowledge, subjects are capable of discerning a gap between a belief and the truth. The original belief might even be classified—albeit in retrospect—as the result of an illusion. But what the evidence also shows is that treating t-belief as incidental leads us to overlook just how deeply ingrained is the tendency to aim away from the truth. If we are compelled to confront our epistemic frailty, in narrowly defined contexts, and just for the nonce, we might be able to respond in accord with TTV. But what TTV omits is the difficulty of accomplishing such a belief revision, the transience of such a revision, and an understanding of why TTV-effects are both difficult and transient. In a word, aiming at the truth can be a very unnatural act.

That TTV cognitions do not come naturally might be the result of their being detrimental in several aspects of our lives. Positive illusions can lead to higher

motivation, greater persistence, and increased likelihood of success (Armor & Taylor, 2003; Taylor & Brown, 1988, 1994; Taylor & Gollwitzer, 1995)—all characteristics that can contribute to the cultivation of self-respect. Athletes, dancers, and soldiers with conviction are more likely to succeed than are those who lack conviction—albeit not nearly so likely as they believe. Positive illusions can also promote use of efficient and rapid problem-solving strategies. There is even evidence to suggest that positive illusions as regards one’s children or one’s partner are critical to successful parenting and to long-term relationships (Barelds-Dijkstra & Barelds, 2008; Wenger & Fowers, 2008).

Lionel Tiger (1999: 617) has argued along similar lines that “moderate” optimism is essential to overcoming our cognitive ability “to generate endlessly discouraging predictions of the pitfalls of any action.” He argues (1999: 615) that we are endowed with a “cognitive override . . . a moderate design defect of pure reason,” something that overrides “cognitive literalness,” that “biases the odds in favor of action” (1999: 619). Among many other supporting observations, he records that recent examination of the dentition of pre-hominids reveals that 3.5 million years ago our East African savannah ancestors were eating large amounts of meat when prey animals were hard to catch. Concerning this point he observes that those who woke up thinking “‘What a great day to catch an ungulate’ would enjoy an advantage over fellow citizens who turned off the alarm and rolled over to sleep straight through the prey’s spurt or morning activity” (1999: 616, also see 1985).

But more is involved than just enhanced performance. Positive illusions can be adaptive for psychological health and well-being. Some evidence suggests (Alloy, 1995; Alloy & Abramson, 1988; Alloy & Ahrens, 1987; Taylor & Brown, 1988, 1994) that there is a group of people who accept both the good and bad about themselves: they remember both good and bad self-relevant information with equal frequency; their evaluations of self and others are congruent; their self-appraisals more nearly coincide with appraisals produced by impartial observers, and so forth. The group of people in question are those “who are low in self-esteem, moderately depressed, or both.”¹² It is sometimes said of these people that their beliefs bespeak a “depressive realism” (Alloy, 1995; Alloy & Ahrens, 1987). When well-adjusted people process self-relevant information, they tend to be biased and partial; those who are dysphoric tend to be unbiased and balanced. Perhaps the single most distinctive finding in this regard is that depressed subjects, dramatically unlike those who are not depressed,

“are consistently accurate judges of their control over events” (Alloy & Abramson, 2007: 242).

The claim is not that positive illusions are a necessary condition for mental health; rather, it is that these illusions can promote mental health (Taylor & Brown, 1994: 25). But that is not all. Positive illusions also seem to be protective of physical health (Taylor, Kemeny, Reed, Bower, & Gruenewald, 2000). For example, studies of AIDS patients reveal that those who believe they can control the disease and prevent its recurrence, those who do not “realistically” accept or appraise their condition, both exhibit a longer asymptomatic period and live longer, by an average of nine months.¹³ Studies of breast cancer and of AIDS patients also show that even the eventual disconfirmation of erroneous beliefs does not have harmful consequences. Moreover, what is true of the sick is also true of the healthy (Taylor, Lerner, Sherman, Sage, & McDowell, 2003): those with positive illusions, while undergoing stressful tests in a laboratory setting, exhibit lower cardiovascular responses, quicker recovery, and lower baseline cortisol levels.

Strategically aiming away from the truth contributes to enhanced performance, a sense of well-being, and better physical health. Significantly, the findings concerning physical and mental health are further confirmed insofar as they dovetail with studies of placebo and nocebo effectiveness. These carefully studied beliefs, when coupled with the studies of positive illusion, are redolent of the example sketched above, in a way that suggests an explanation for the durability of superstition.

A placebo effect is that which follows from the administration of a pharmacologically inert substance or physiologically inactive treatment¹⁴ that is coupled with the verbal suggestion of clinical benefit. Nocebo effects also follow upon administration of an inert treatment, differing from placebos in that they are accompanied by suggestion of clinical harm (Benedetti, 2008; Diederich & Goetz, 2008; Oken, 2008; Zubieta & Stohler, 2009).¹⁵ Nocebos, in that their effects are adverse, bear more direct resemblance to the alleged consequences of ignoring superstitions.¹⁶ What matters though is that the nocebo or placebo, despite being inert, by virtue of engaging a person’s belief—in a manner that aims away from the truth—is able to bring about a measurable physiological outcome, salubrious or noxious (Benedetti, 2008: 36, 48).

Placebos have been demonstrated to have salubrious effects in the treatment of many conditions, e.g. pain, swelling, addiction, cardiovascular and respiratory problems, peptic ulcers, depression, anxiety, cancer, and Parkinson’s disease

(Benedetti, 2008; Evans, 2004). Some of the mechanisms¹⁷ whereby belief is able to effect these changes include: the release of endogenous opioids or dopamine, the inhibition of serotonin uptake, the reduction of β -adrenergic heart activity, as well as the conditioning of immune receptors like lymphocytes and hormones like cortisol. Further, differentiation among types of placebo effectiveness are being teased apart: for example, conscious expectation seems to play a greater role in alleviating pain and enhancing motor performance, whereas classical conditioning can be sufficient to trigger immune and hormonal responses (Benedetti, 2008: 42; Nieme, 2009).

Once again recall that, according to Velleman, when we discern a gap between a belief and the truth, the belief becomes unsettled and starts to change. Studies of placebos, however, reveal a dissociation between different forms of belief regulation: one results from conscious expectation, the other, from classical conditioning. A natural explanation of superstition susceptibility now suggests itself. A person who, sincerely and for good reason, denies holding the belief that 13 is unlucky, might, due to analogues of classical conditioning that occur in everyday life,¹⁸ come to behave in such a way that can best be explained by the belief that 13 is unlucky. Even if acting in that way contravenes professed beliefs, it is not obvious these non-evidential considerations (those regulated by classical conditioning) can only be influential if unacknowledged. Although this might strike some as absurd, perhaps the apparent absurdity merely reflects a design compromise that has been achieved during our evolutionary development.¹⁹

Placebo beliefs are like positive illusions in that both are false. But placebos are false in a distinctive way. To illustrate this point, compare placebo effectiveness with positive illusions that cause people to be overconfident in the extent and effectiveness of their control. What they are right about is in believing that they exercise some control; they are wrong, however, in their assessment of how much control they have. A person who asserts that he has the ability to hit a 450 foot home run might be wrong by a degree that is easily calculable. But a person who asserts that by drinking a particular potion (perhaps a mix of tap water, sugar, and food coloring) his peptic ulcer will be cured is completely wrong. There is nothing in the potion that will contribute to his cure; it is pharmacologically inert. Nevertheless, effective brain mechanisms can in this way be set in motion.

Today placebo effects are often triggered by the presence of doctors, medications, needles, even just the smell of a clinic, all things that are highly correlated with effective treatment. But this could not have been the case within which placebo

mechanisms evolved. And, for both modern and antediluvian placebos, we know for a fact that any correlations which might obtain are non-causal.

Consider the candidate mechanisms cited above: the release of opioids or dopamine, the conditioning of immune receptors, the inhibition of serotonin uptake, or the reduction of β -adrenergic heart activity. Presumably they are activated by means of some form of mind-body “lingua franca” (Humphrey, 2004; see also Beauregard, 2007: 233), the psychological side necessarily involving false beliefs. Also note that modern medicine is scarcely more than a century old.²⁰ we are not long past the days of blood-letting and ignorance of microscopic organisms. The mind-body lingua franca apparently evolved in an environment under which the input constraints definitive of (those) beliefs could not have been yielding true beliefs. After all, lacking alternatives, systematic examination of shamanic beliefs, rituals, and incantations, including careful consideration of instances of failure, would hardly have been worth the effort. And morbid acceptance of death and disease was likely no more helpful to individual or group esprit during the Pleistocene than it is today. So it is for good reason that the placebo effect is, as Bakan (1985: 212-213) has written, demonstrated “precisely in cases in which expectancy is falsely grounded.”

To say that these beliefs aim away from the truth is not to say that they also aim away from “instrumental success.” But recall that Velleman insists such goals are “second-order.” The contention here, by contrast, is that belief regulatory systems evolved at a time when aiming for the truth, in some areas of life, would have been as pointless as counseling Aristotle to devote more attention to the brain—millennia before the development of neuroscience. In this area of life, for good reason, instrumental success supersedes truth-conduciveness.

Reflection on the modern world suggests that communities maintain repositories of false beliefs and humans retain wells of gullibility that can be drawn upon during times of social turmoil or personal crisis. To cite just one example, Pascal Boyer (2000: 99-100, 105-106) has found that when dealing with religion we are naturally inclined to be gullible as concerns the “odd” or the “unfamiliar.” Paradigmatic of these are spirits who are represented as intentional agents, but agents whose physical properties violate the physical qualities of embodied agents. Not only are these violations not taken as evidence that the entities aren’t real, instead, “it is precisely insofar as a certain situation violates intuitive principles and is taken as real that it may become particularly salient” (Boyer, 2000: 101). In effect it appears that

appropriately structured systems of false belief remain available within society, accessible to all, and ready when needed.²¹

A third set of beliefs that aim from the truth are referred to as self-deception (Mele, 1997: 92, 2001: 4). These too are extremely common: paradigmatic examples include people who believe in their spouse's fidelity, their likelihood of recovery from illness, or their child's avoidance of illicit drugs, despite the availability of evidence so compelling that were it about the spouse, the illness, or the child of someone else, their confidence would surely be shaken. Alfred R. Mele (1997: 93-94) has noted that these self-deceptions prime other cognitive mechanisms, such that they then contribute to the production of yet more false beliefs, mechanisms that include information salience, the availability heuristic, confirmation bias, and our tendency to search for causal explanations. If Mele is correct, then one false belief can lead to a concatenation of further false beliefs. According to Robert Trivers (2000: 125) some self-deceptions function as do positive illusions. But he (1985, 2000) also suggests that self-deception has a unique role: it was favored by natural selection because it enhances our ability to deceive others. The idea is that if we first deceive self (e.g. a politician who says to his constituents, "I feel your pain"), then the autonomic nervous system changes that might indicate falsehood to others would not be manifest. Creatures capable of self-deception could then reap certain Machiavellian rewards.²²

There is abundant evidence that sometimes beliefs are regulated so to aim away from the truth. What's more, the mechanisms engaged in production of false belief are difficult to override. Consider, for example, that rejection of a superstition might have been regulated in truth-conducive ways. And because the rejection is so thorough, it should not be subject to truth-conducive revision or extinction. But, prior to rejection, if one has been classically conditioned, in that the superstition was learned early in life and under the proper conditions, despite having later been extinguished, self-control will still be required to resist its influence. According to the ego-depletion hypothesis, self-control is like muscle strength (Muraven&Baumeister, 2000). It is a limited resource that can be exhausted by excessive demands and, once depleted (e.g. as indicated by low levels of blood glucose), recovery is slow (Galliot et al., 2007). Accordingly, we should not be surprised to discover that when the ego is depleted, as can happen when one is under great stress, people might behave in ways that contravene professed beliefs. And if one is trained to recognize the indicators of depletion, it would not be surprising to find that they are capable of acknowledging

the role non-evidential considerations— e.g. low-levels of blood glucose—play in determining what is believed.

When considering whether beliefs might aim at something other than the truth, recall that Velleman invokes a basketball analogy. Regardless of whether this analogy can be cogently applied to beliefs of any kind, it certainly does not fit here. By that analogy, within the game, one plays to win (the first-order aim), because doing so will ensure salary increase (the second-order aim, the ultimate aim). To take the case of placebos as an example, clearly their ultimate—the second-order—aim is to be restored to good health. But to realize that ultimate aim, one cannot adopt a first-order goal of aiming at the truth; to do so would be self-defeating. On the court—in the game of life—it is not the case that these beliefs are truth-directed. In these contexts, truth-directedness and instrumental success are at odds with one another.

V. The Distinctive Character of Tertullian Beliefs

What is most distinctive about the cluster of beliefs described above is that they aim away from the truth. I have dubbed them Tertullian beliefs, or t-beliefs.²³ Tertullian seems a proper eponym because it is (apocryphally) said that he proclaimed: “I believe because it is absurd.”²⁴ What he (Tertullian, 2010) actually wrote was, “certum est, quia impossibile:” that is, “it is certain, because it is impossible.” The idea of believing because it is “absurd” or “impossible,” though a bit hyperbolic, evinces the deliberateness of aiming away. What matters is not that the beliefs are false. What matters is that they seem calibrated to be false, in a certain way, and to a certain degree.²⁵ The effectiveness of a positive illusion, a placebo, or self-deception depends upon aiming, in just the right way, a calibration which seems achievable only as the result of design.

As a first approximation the metaphor “direction-of-fit” can be employed to capture the distinction between t-beliefs and TTV beliefs (Searle, 2001: 37-38, 257). TTV beliefs exhibit a mind- to-world direction of fit; t-beliefs, on the other hand, exhibit a world-to-mind direction of fit. Mind-to-world direction-of-fit implies that it is the purpose of TTV beliefs to change so that they match the world. Ordinarily world-to-mind direction of fit— representation not of how things are, but of how we would like things to be—is used to characterize the attitude “desire.” Of desire it can be said that its purpose is to change the world to match its content.

T-beliefs, like desire, aim to change the world.²⁶ In this respect, they are unlike TTV beliefs. Nevertheless, they are asserted in such a way as to imply that they represent how things are—13 is unlucky, I am better-than-average, the waters of

Lourdes have curative powers, or I am sincere. T-beliefs are not regarded as true in a fanciful, polemical, or heuristic way. The spirit in which the propositional object is regarded as true is serious. T-beliefs require a blurring of the usual belief-desire distinction. Direction-of-fit can help elucidate this relationship, but it remains just a metaphor (cf. Sobel&Copp, 2001). Fortunately direction-of-fit can be further explicated in terms of causal connectedness. Consider again self-confidence, self-healing, and self-deception. The beliefs associated with these phenomena are generated by mechanisms that aim away from the truth. Not only that, like desire they conspire to change the world to match their content. When they succeed, it is not by accident. Rather their success results from their being about the same part of the world (the body) that they inhabit, either intra-cranially or inter-personally. Positive illusions and placebos can be effective intra-cranially via the appropriate mind-body lingua franca, and self-deceptions can be effective by shutting down autonomic reactions that would otherwise be detectable to those one wants to persuade. This corner of the world—the intra-cranial and the interpersonal— is, so to speak, within striking range of belief. There is a causal link between these beliefs and that portion of the world that they target.²⁷

Note that t-belief is not reducible to desire. People who merely “want” to be better-than-average, to recover good health, or to be inter-personally successful don’t succeed in the way that those who t-believe do. A clear distinction between mere wanting and believing remains: the moderately depressed want to perform at a higher level they just don’t believe that they will. And most who fall ill want to recover. But only t-belief improves chances of recovery by means of the intra-cranial causal nexus. Simple desire, mere wanting, doesn’t cut it.

Another way to approach this distinctive blend of belief and desire is to note that it can help to diminish the puzzlement of a philosophical curio, Moore’s Paradox. Consider that denouncing a superstition but being influenced to act in accord with that superstition seems rather like an instance of “p but I don’t believe that p.” In other words, it is suggestive of what has come to be called Moore’s Paradox, which is just a paradox in the informal sense for “p” and “I don’t believe that p” might both be true. Nevertheless, since asserting “p” seems to imply the belief that p, this is typically regarded as an utterance of a type that I cannot sensibly assert of myself.²⁸ In the superstition case we have apparently contradictory expressions, both the assertion “I don’t believe that p” and behavior which seems best explainable by attributing the

belief “p.” Typically it is claimed that one could not self-ascribe both.²⁹ But when one is aware of what is implied by one’s behavior, such self-ascription is possible.

Jeanette Kennett and Cordelia Fine (2008: 176-177) have found that psychopaths and sociopathic delinquents produce many statements that are “Moorean paradoxical” (cf. Joyce, 2007: 51-57). As a typical example, consider: “John is an honest person. Of course, he has been involved in some shady deals!” As with Moorean paradoxes generally, when treated as a whole, the statement seems to make no sense. Kennett and Fine treat this paradox as a measure by which to determine whether the psychopaths or sociopaths grasp what is implied by evaluative terms.

What I am suggesting is that healthy people are capable of Moorean paradoxical expressions in that their professions of belief are contravened by behaviors whose implications are recognizable to the subject. This tension between what one professes and how one behaves reflects a design compromise, one which for most people is salubrious. If this view is correct, we can reasonably expect that those who are mildly depressed, should be more sensitive to the implications of Moore’s paradox; therefore, they would be less likely to behave in ways that contradict their professed beliefs.³⁰

One might wonder why this aspect of belief has yet to be duly recognized. After all the relevant scientific studies can now be traced back to well over two decades. Perhaps it is that one of the institutions which consistently gives these beliefs pride of place, religion, is not taken seriously.³¹ Perhaps as well we live in a world with so many dangerous false beliefs that we fail to appreciate non-TTV forms of belief regulation (Bennett & Hacker, 2003: 172-174). A further factor that causes neglect of t-belief might derive from an under-appreciation of what Wallace Arthur (2004) calls “internal adaptations.”

Arthur (2004: 117-127) points out that “ecological” adaptations, adaptations to the (external) physical environment, receive most attention in biology; internal adaptations or “coadaptations,” adaptations among body parts, tend to be neglected. An example of the former is the adaptation of forest flies to higher ambient temperatures (Arthur, 2004: 122-123): flies must struggle to stave off desiccation. The hotter it gets, the faster they lose water. Because the larger one is, the smaller one’s surface area is relative to volume, and because water loss occurs at the body’s surface, in a dry, hot environment being bigger is better. In accord with selective pressure, the average body size of the fly population will increase. Because the fitness difference is clearly produced by the external environment, this counts as an external adaptation.

But suppose that along with the difference in body size, these flies also differ in the way their wings are connected to their thorax. Suppose as well that this variation slightly affects their ability to fly. Under such circumstances, the population will evolve toward better integrated joints. Here though selection is unrelated to the forest's change in ambient temperature; it isn't even related to the forest. Although flight occurs in environments, good flying ability is generally advantageous for flies, no matter what environment they inhabit. Accordingly, these fitness differences are "quasi- environment-independent." Internal selection, in an important sense, "travels with the organism wherever it goes."³²

Relating this distinction to t-belief, we might say that most philosophical attention to belief has concerned "ecological adaptations." Understandable though this might be, the "internal" environment is also part of that to which we must adapt.³³ And because t-belief is so critical to internal adaptations, it warrants more attention than it receives from within the TTV conceptual framework.

W. V. O. Quine (1994: 66) famously wrote: "Creatures inveterately wrong in their inductions have a pathetic but praise-worthy tendency to die before reproducing their kind." This clever turn-of-phrase strikes many as necessarily true. But it misleads. Sometimes we are wrong for good reason. And, to grasp what counts as a good reason, we should attend to internal adaptations.³⁴ A balance must be struck between the external and the internal, between TTV and t-belief. Accuracy of inductions concerning the external world is not enough. Neglect of internal adaptations can also lead to pathetic but praise-worthy ends.

VI. Wood's Procedural Principle and T-Beliefs

Recall that Wood (2008: 13, fn. 8) presupposes that no one can stably hold a belief they know to be false. But there appear to be counter-examples to this claim. In the anecdotal cases, people behave in accord with superstitions and purchase vanity products or resort to miracle cures, even when they can fairly be said to know that recourse to these strategies or products is grounded in false beliefs.

As regards empirical studies of t-beliefs, even when subjects are presented with explicit description of cognitive biases immediately after acting in accord with those biases, they exhibit no evidence of belief instability. If the notion of "instability" is unpacked in the way proposed by Velleman—i.e. gaps between belief and truth are reconciled by either adding new beliefs or changing those originally held—it seems that at most the instability of believing is narrowly circumscribed. Only if compelled to confront evidence in constrained experimental settings might one evince the

predicted adjustments. And even these meager findings might not be ecologically valid. In sum, there is no evidence, independent of claims based upon contentious introspective reports, that belief instability is a natural disposition.

As regards the ethics of belief, Wood (2002: 38-40) acknowledges the possibility of exceptions to the procedural principle. But he regards criticisms of it that are based upon this possibility as “cheap” and “wrongheaded.” Even should a person determine that the principle need be violated, Wood counsels that the person should “feel squeamish and conflicted.” Wood’s (2002: 33) harsh judgment in this regard is motivated

by his belief that violations of the procedural principle are “shameful in something of the same way that telling lies is shameful.” When we believe that which is “comfortable to believe,” we show contempt for self and perform a disservice to others. That is, we fail to respect ourselves as rational beings and we deprive others of honest evaluations that they might need.

Doubtless Wood (2008: 13) is correct that one need not look long or far to find innumerable examples of “shameless evasions.” But if we apply a principle of psychological realism to our moral theories,³⁵ then there are reasons to be dubious of the procedural principle. First, there is good reason to believe that gaps between belief and truth do not necessarily precipitate instability. We seem to be designed in such a way as to allow for these inconsistencies, without the untoward spill-over effects that concern Wood. A stable compromise has been forged between internal and external adaptations. The more closely one examines instability claims, the more they seem to be artifact derived from unwarranted philosophical expectations of consistency.

Second, if we were to feel squeamish and conflicted each time we acted in accord with a positive illusion or with the distribution of a placebo, the benefits of illusions and placebos would not be attainable. As regards self, for the positive illusions to contribute to our well-being (and not, say, exacerbate depressive realism), we should not feel squeamish or conflicted. As regards our treatment of others, for the placebo to be effective, likewise, we should not feel squeamish or conflicted, as these would be evident to the patient.³⁶ And if the benefits of t-belief were not forthcoming, the results would include diminished health, performance, motivation, and well-being.

In addition to Wood’s skepticism that there are legitimate exceptions to the procedural principle,³⁷ he believes that violations are essentially “corrupting” (2002: 36). He believes that people are inclined to take unacceptable liberties, allowing for both unrestrained rationalizations of personal behavior and dishonesty in public

discourse. But t-beliefs seem to be legitimate exceptions, and there is no empirical evidence that the tendency to act in accord with t-beliefs leads to the corrupting tendencies that are the object of Wood's concern. It might be the case that Wood is correct in his assessment of other beliefs. But it is not difficult to conceive of people who hold positive illusions, self deceive in the standard circumstances, and react to placebos in ways that enable them to be effective, while not allowing for these breaches of the procedural principle to adversely affect other aspects of their lives or of public discourse. T-beliefs seem designed so as to be insulated from the rest of our beliefs.

What Wood's advocacy of an uncompromising adherence to the procedural principle requires is evidence of a particular sort. For example, if it turns out to be the case that depressive realists are less inclined to corruption and more respectful of self than are the majority of people, then Wood's views could be said to be rightly affirmed. But there is no evidence of this sort; none, whatsoever.

It might be said that Wood's view reflects a strictly normative position, and that empirical evidence, be it anecdotal or scientific, is of no relevance. But, just as a matter of fact, Wood justifies his uncompromising position by making specific empirical claims about the nature of belief as well as about the tendencies of people who fail to act in accord with the procedural principle. To show that these empirical assumptions are dubious as regards t-belief then is to weaken support for Wood's version of this principle.

Furthermore, the evidence suggests that, contrary to what Wood maintains, t-beliefs might be critical to—rather than detrimental to—the maintenance of self-respect. What is neglected is the compromise between internal and external adaptations, as well as the causal role that t-beliefs can play. TTV tends to treat t-beliefs as incidental; Wood takes this view a step further and treats them as “pernicious.” But when properly calibrated, they can help alleviate depression, improve health, enhance motivation, and improve performance. Depression, ill health, indolence, and failure are not contributors to self-respect. They are obstacles. What both anecdotal and empirical evidence suggest is that an appropriate dose of the right kind of false beliefs might be a necessary condition for the development of self-respect. Sometimes it pays to be Moorean paradoxical.

Recall Tiger's speculation concerning the lot of a pre-hominid who lacked the capacity for t-believing. It is simply too easy “to generate endlessly discouraging predictions of the pitfalls of any action” (Tiger, 1999: 617). We seem to need an

antidote to “cognitive literalness,” something that moves us to action. T-beliefs, in right measure, just are that antidote. Without them we are less inclined to taking action in a whole host of ways that are essential for self-respect.

Wood (2008: 19) emphasizes that self-respect requires the apportioning of belief strictly in accord with the evidence. But those who best adhere to this requirement as regards beliefs about self tend to have low self-esteem (Alloy, 1995; Alloy & Ahrens, 1987). The claim advanced here is that a certain measure of self-esteem is a precondition for self-respect. Those who are without t-beliefs seem to lack the minimum esprit necessary for the maintenance of that which Wood values so highly.

VII. Conclusion

Wood (2002: 8, 2008: 9) emphasizes that we are responsible for the processes of belief formation and maintenance. Just as we would be blameworthy for killing someone in a drunken rage, so too we are blameworthy for acting in accord with beliefs that are not properly formed or maintained. In the former case, we should have known not to get drunk. In the same way, we behave irresponsibly when we allow cognitive biases to lead away from the truth. We are obliged to be proactive.

Might Wood have a point here? Perhaps the way things are with beliefs is blinding us from the way things could be. Perhaps we, individually and collectively, need to be weaned from t-beliefs. And perhaps this would be a good thing. But the formal investigations of Pronin (2007) suggest that weaning is not an option.

Less formally, it seems to be the case that when progress is made toward reducing the effects of cognitive biasing on one front, those biases reemerge on another. Above I noted that, in France, as the number of Roman Catholic clergy decrease in numbers, the number of professional astrologers increase. A Conservation of Credulity Principle seems to be in effect.

Whether or not individuals or societies can be weaned from t-belief in such a way as to manage proper alignment with the procedural principle is an empirical issue. Whether or not we should be weaned is an ethical issue. For Wood, it would seem, the two become relevant to one another when we assess the cost of attempts at weaning. If success brings about enhanced self-respect and no collateral, corrupting effects, then it is a good. If it brings about diminished self-respect and an increase in corrupting effects, then it is not. If the latter, then even on Wood’s terms, the procedural principle should be compromised.

ii. Issues at the intersection of ethics, evolution, and

neuroscience

There was a time when ethicists did not concern themselves with the natural sciences. Even in the year 2010 it might be accurate to say that the overwhelming majority of ethicists, other than those who are concerned to prescribe proper conduct for scientific practice, pay little attention to natural science. It has often been said that science primarily concerns itself with what is the case, while ethics concerns what should be the case. If this is true, then ethics research can or should be conducted without caring too much about what the sciences say.¹

But in recent decades, among some philosophers² the idea that the sciences—in particular evolutionary biology and the cognitive neurosciences—have much to contribute to our understanding of ethics has been gaining traction. This is not to say that such an approach to research in ethics is entirely new, for clearly that is not the case. Many who now treat ethics as a field of study that is done best when animated by reflection on the findings of contemporary science are extending ideas that were foreshadowed in the works of Aristotle, Mencius, and David Hume, among others. Although aspects of the conceptual framework have been in place for centuries, only in the 19th and 20th centuries, with the development of evolutionary biology, and in the 20th and 21st centuries, with the development of cognitive neuroscience, have these ideas been refined through synthesis with systematic empirical investigations.

If science concerns what is the case and ethics concerns what should be the case, then might those who argue that the two disciplines should remain distinct be correct? Among those who promote some version of naturalized ethics, unanimity of response has not yet been achieved. But the positions staked out by Neil Levy and Owen Flanagan are representative of a general tenor expressed in the works of those ethicists who engage evolutionary biology and cognitive neuroscience.

Neil Levy (2007), who has authored the essay “The Prospects for Evolutionary Ethics Today” for this issue of *EurAmerica*, emphasizes that the emerging “neuroscience of ethics” might reshape our understanding of certain fundamental, ethical concepts—e.g. agency, free will, intuition, and rationality. Were this to be the case, the ramifications for theorizing over ethical matters would be substantial. Nevertheless, Levy’s point of departure is not altogether unfamiliar to traditional ethicists: he is in sympathy with Rawls’s (1971) view that in moral inquiry we seek a reflective equilibrium among our intuitions and our moral theories.

Levy, however, differs from many traditional ethicists in several respects: first, although he believes that our intuitions can have justificatory force, he does not

regard them as sacrosanct. Second, he regards many questions that are pivotal to ethical theorizing as straightforwardly empirical (e.g. whether self-interest is our primary motivation when rendering moral judgments). Third, he takes seriously the idea that aspects of the external world (anything from a sextant, to a sacred scripture, to a member of our social cohort) play an essential role in human cognitive activity.³ One significant implication of the view that cognition is extremely dependent upon the external environment and multifarious props is that morality should be treated as a social enterprise, an enterprise that takes heed of expert counsel and that strives for overall consistency.

Owen Flanagan (2002, 2007) who, in collaboration with David Barack, has authored the essay “Neuroexistentialism” for this issue, echoes Aristotle in exhorting us to conceive of ethics as systematic inquiry into the conditions necessary for leading a good life, conditions that promote flourishing. In other words, Flanagan’s treatment of ethics is more inclusive than is the work of some other ethicists: he is concerned both with what is moral and with what makes life meaningful. But though these conjoined concerns mark his work as distinctive, they do not mark his approach as unconventional. What is more likely to cause consternation, at least in some quarters, is his treatment of ethics as a kind of applied science. More specifically, he treats ethics as being like ecology: just as we might seek to identify the conditions that permit various natural systems (e.g. the oak-hickory forests of the Ozark Mountains or the cypress forests of Mount Ali) to flourish, so too we might seek to identify the conditions under which humans can best flourish. Hence, ethics is best regarded as a kind of “human ecology.”

Although, on this construal, ethics is empirical, Flanagan does qualify this claim somewhat. First, he does not anticipate that ethics will turn out to be like physics, allowing for the derivation of causal generalizations from general laws. On the contrary, many among the significant generalizations that are to be found will be just as they are in ecology, singular and local. Second, ethics as human ecology is a normative science, in that it goes beyond description, explanation and prediction; it includes inquiry directed at discovering the conditions which must be satisfied in order to attain certain ends. If you want to build a skyscraper that won’t collapse in an earthquake, you should satisfy certain conditions. Likewise, if you want to foster a society or a human being that flourishes, you should satisfy certain conditions.

How does one determine proper goals, and what counts as flourishing? Fortunately ethical inquiry that engages science need not ignore the centuries of

wisdom that accumulated prior to the advent of human ecology. For example, like Levy, Flanagan too draws upon Rawls (1971), who in turn draws upon ancient wisdom. Rawls observes that the “Aristotelian Principle” can serve as a guide to flourishing: human beings enjoy the exercise of capabilities, whether innate or trained, and the more complex the better. Of course from the perspective of Flanagan’s human ecology, this can only be treated as a hypothesis about human psychology.

Might it turn out to be the case that people and environments differ so substantially that we inadvertently open the door to a pernicious form of ethical relativism? Since ethics as it is considered here is an empirical inquiry, the possibility cannot be dismissed out of hand. But certain vices and virtues appear to be recognized universally, recognized in all human habitats. These seem to reflect a shared body of fundamental intuitions, including intuitions pertaining to the just treatment of those who are neither kith nor kin.

Here too I believe Levy’s and Flanagan’s views dovetail. We need not worry excessively about the possibility of pernicious relativism, because sometimes our intuitions can rightly be said to have justificatory force. Intuitions concerning justice might well be one of these. Furthermore, Flanagan also endorses a view consonant with Levy’s, that morality should be treated as a social enterprise, an enterprise that takes heed of expert counsel and that strives for overall consistency. On Flanagan’s account, a racist, a xenophobic, or a misogynist might feel happy, but it is likely that through dialogue and through the discoveries of experts (including evolutionary biologists and neuroscientists) that these attitudes will be found—as a matter of fact—not to promote environments in which people flourish.

“The Prospect for Evolutionary Ethics Today,” by Levy, is an attempt to allay worries that acknowledging morality’s evolutionary origins might imply abandonment of integral notions of morality. He traces the erroneous reasoning that has given rise to the worries expressed in the work of some contemporary evolutionary ethicists to the dispute between Thomas Huxley and Herbert Spencer as regards how best to understand Charles Darwin’s ideas on natural selection, particularly as these relate to the proto-morality of our evolutionary ancestors. While not denying that our moral sentiments are the product of evolution, Huxley argued that ethics is—in some important respects— independent of our biological nature. After all, according to Huxley, our immoral sentiments are also the product of evolution, so why should we privilege one over the other? Accordingly, he held that

“good” doesn’t mean “adaptive” and that morality should be designed so to stand in opposition to evolutionary processes.

Spencer, who coined the phrase “the survival of the fittest,” thought otherwise. For him, “good” just means “highly evolved.” This meta-ethical position has significant normative implications. For example, Spencer counseled against organized charity, because it would ameliorate the suffering of those who are genetically destined to fail. Eugenics, on the other hand, was endorsed by the Social Darwinists inspired by Spencer, for that the “highly evolved” should survive—even if at the expense of those “less highly evolved”—is taken to be a good.

Levy argues that morality, properly understood, implies that we should side with Huxley: that is, we should sometimes, in some respects, combat natural selection. All parties to this dispute can agree that certain raw materials—e.g. altruistic dispositions—are evolutionary products. What the “neo-Spencerian” needs though is evidence and argument to show that morality is to be identified with those raw materials, the constituents of proto-morality. Levy’s concern is not that identifying proto-morality with morality would be to run afoul of the naturalistic fallacy. His concern is that the analysis whereby one might determine the two to be identical simply fails. For example, any analysis that would conclude that “good” is equivalent to “highly evolved” would fail, because it would imply that certain propositions which we hold dear—e.g. xenophobia is bad and charity, good—are false.

Why not then just conclude so much the worse for those propositions we hold dear? The reason is that our innate dispositions often conflict with one another; even our evolved altruistic intuitions are discordant. Although we have been endowed with a partial sensitivity to the needs and interests of others, it is a sensitivity that is geared principally to self-interest. But at the same time we have been endowed with the belief that our moral sensitivity should not be predominantly self-interested. Because these dispositions are at odds with one another, they can only serve as a starting point. Rationality is needed to trim and refine them such that we might approach a reflective equilibrium.

“Neuroexistentialism,” by Flanagan and Barack, focuses on one of the issues that makes achievement of reflective equilibrium so difficult—the clash between scientific and humanistic images of persons. Like previous existentialisms, neuroexistentialism is a response to a diminished self-image. In this instance, the third wave of existentialism, neuroscience has added evidence that makes Darwin’s message especially vivid, making it all but impossible to ignore. That message is that we are

animals; the mind is the brain; and, that we are one kind of fully material creature living in a fully material world. The worry, to put it baldly, is whether we can flourish, given that we know ourselves to be nothing over and above social, embodied creatures, creatures with an evolved capacity for rationality.

According to Flanagan and Barack, proponents of Darwinian views often fail to see that opponents are correct about a matter of vital import: if Darwinian views are correct, then what people are justified in believing conflicts with antecedently held views of who or what we are. Because the humanistic view does not mesh well with the scientific image, we can find ourselves cast adrift, in an anchorless search for meaning of the sort that characterizes all existentialisms. It is then no wonder that advocates of creationism and intelligent design are taken so seriously in the United States.

Flanagan and Barack distinguish their concern from that which David Chalmers (1996) has dubbed the “hard problem” of understanding how it is that consciousness is realized in the electro-chemical activity of brains. If we allow that the cognitive neurosciences will provide us with an answer to that how-question, we are still left with a “really hard problem.” Given that everything about us, including consciousness, just is part of the natural world, can anything that is both uplifting and true be said about the meaning of life. Unlike the “hard problem,” the “really hard problem” is not a purely scientific question. It concerns a philosophical attitude: in view of the fact that we are evanescent members of a species that will one day become extinct, how should we regard ourselves?

One form of descriptive-normative inquiry that might help to quell neuroexistentialist anxiety is “eudaimonics”—the study of those conditions which promote flourishing or fulfillment. Fortunately, eudaimonic inquiry need not start from scratch. Both modern science and works of philosophy that have accumulated over the ages, provide many resources that can be drawn upon in designing suitable responses, responses that do not resort to the supernatural, the theological, or the transcendental.

Flanagan and Barack conclude by raising a worry: some findings within the cognitive neurosciences seem to suggest that positive illusions might importantly contribute to eudaimonia. Were this the case though, we would need to choose between believing what is true and living a life that enables us to flourish. Flanagan and Barack, however, express the hope that the need for positive illusions is not intrinsic to human nature.

But “The Ethics of False Belief,” by Lane, takes seriously the idea that positive illusions, as well as other forms of false belief, might be intrinsic to human nature. He considers both anecdotal and scientific evidence which suggests that this might be so. He proceeds then to argue that some of our beliefs might be the result of an evolutionary compromise between internal and external adaptations. Not only should we believe what is true, if we are to survive well in this world, but we should also, sometimes, strategically, believe what is not true. Believing what is not true is a form of internal adaptation. It is an adaptation to being the kind of animal that knows it is a frail and mortal member of a species destined for eventual extinction.

The essays collected here presuppose that we are evolved creatures whose minds depend (in one way or another) upon our brains. They also share a commitment to the view that contemporary ethics is done well when it is animated by the findings of evolutionary biology and the cognitive neurosciences. But no one among these authors would claim that a consensus has already been achieved for how best to conduct research of a neuroethical sort. Nevertheless, like philosophers of any era, at least those philosophers who continue to be taken seriously in the 21st century, they draw upon the resources that are available to them in the era within which they work. In this era it would be foolish to neglect what we are learning about our evolutionary origins or to ignore the discoveries of neuroscience.

1 What I here refer to as the procedural principle is also known as “Clifford’s Principle” (Wood, 2002) or the “evidentialist principle” (Wood, 2008: 10). I use “procedural” rather than “Clifford’s” to emphasize that my concern is with Wood’s version, and I refrain from using “evidentialist,” because this term is more often employed by the principle’s critics than by its advocates.

2 Although the expression is sometimes used metaphorically, it can also be used literally, as when one is speaking of the aims of people who form beliefs or of design mechanisms that constrain the regulation of beliefs.

3 Belief and imagination can be combined though, as in cases of metaphorical belief (McGinn, 2004: 134). A simple example would be employing a simile to express

4 one’s belief, such as “the sky is like the ocean.” Although Wood speaks more of evidence (empirical, a priori, etc.) than of truth, it is clear that what matters for both is this—responsiveness to evidence and reasoning (2008: 10). He makes the implied conceptual relationship between evidence and truth explicit when he writes that there is “no other responsible guide to what beliefs are true than that which the evidence indicates” (2002: 70-71).

5 Shah and Velleman: “It is an objection to belief that it is false . . . it is a fatal objection, in the sense that if the person who has the belief accepts the objection, he thereby ceases to have the belief, or at least it retreats to subconscious . . .” (2005: 531, fn. 16). See also Williams (2002: 67). The received view of beliefs is that some are conscious, some not. But extreme positions do exist: Searle holds that all beliefs are conscious (1992); Crane (2001: 103-108), that none are.

6 For those readers for whom the number 13 fails to evoke superstitious anxieties, substitute any superstition that does and construct a scenario parallel to the one sketched here; nothing hinges on this particular example. Gazzaniga (2008: 271-272), for example, cites the example of walking quickly past a cemetery at night, even though one doesn't believe in ghosts. The number 13 example is used only because it is familiar to a wide audience and because it has been demonstrated (Scanlon, Luben, Scanlon, & Singleton, 1993) to consistently and

7 significantly affect behavior. Case (2000) and Case, Fitness, Cairns, and Stevenson (2004) have provided some experimental evidence to support the claim that even skeptics readily resort to superstition. For related material see Shermer (2002, 255-313), Talmont-Kaminski (2008), and Vyse (2000). Talmont-Kaminski generalizes from the data 8 to assert that superstition is a basic human trait.

It might be thought that in contexts like this belief should be understood in a Bayesian way, i.e. as the assignment of probabilities to statements. But to do so would mislead, for the subjects express categorical denial. Moreover, on the Bayesian construal, it is perhaps more aptly said that beliefs are just "tentatively" held; therefore, it would be incompatible with TTV. Finally, although probabilities of statements can be applied in certain situations, still it would seem that those situations must then be believed to be of that type by a subject. In other words, Bayesian conceptions seem to presuppose the attribution of non-Bayesian beliefs (Nozick, 1993: 94-99).

9 I am presupposing that few people have the intellectual courage of a John Diamond (2001), who steadfastly refused to yield to superstition or ungrounded claims, even though he was gravely ill. Instead, he devoted his time to his attempt to complete "Snake Oil," his critique of alternative medicine. Also, note that "for the most part intelligence is orthogonal to and independent of belief" (Shermer, 2002: 285), that educational level does not influence susceptibility to superstition (Case, 2000), that maintaining a dubious attitude toward a proposition requires energy (Gilbert, 1993), and that stress and uncertainty incline one to resort to

10superstition (Keinan, 2002).Bausell (2007) provides a detailed survey of complementary and alternative medicine (CAM); a preponderance of evidence shows that nearly all CAMs, including echinacea, are ineffective.

11 For a critical assessment, see Colvin and Block (1994). Some (Heine, 2001: 897-900) have questioned whether positive illusions are universal. The preponderance of evidence (Acker & Duck, 2008; Church et al., 2006), after allowing for some conceptual refinements and methodological tinkering, indicates that they are.

12 What seems to be true of the moderately depressed is not necessarily true of the severely depressed. As regards the latter, findings are equivocal (Alloy & Abramson, 2007; McKendree-Smith & Scogin, 2000).

13 One physiological factor that seems to contribute to the non-realists more robust 14 health is their ability to maintain a higher level of CD4 T helper cells.

A placebo can be any clinical intervention, whether words, gestures, pills, various devices, or surgery. There are even hierarchies of effectiveness: e.g. injections are more effective than pills, and incisions more effective than injections (Evans,

15 2004). Verbal suggestion is not essential; sensory stimuli in an evocative setting (e.g. the sight of a syringe while sitting in a clinic) can be sufficient to elicit the effect. But when considering the effects of placebos, one must be careful to factor out other causes, e.g. spontaneous remission, regression to the mean, and patient biases. Moreover, effectiveness can vary. This variation

might be explainable in terms of functional differences in the mesolimbic dopaminergic pathway (Scott, Stohler, Egnatuk, Wang, Koeppel, & Zubieta, 2007).

16 Perhaps the most famous documented example of a placebo effect is “voodoo death” (Lex, 1977).

PET technology (Mayberg, Sliva, Brannan, Tekell, Mahurin, McGinnis, et al.,

2002) has made it possible to begin teasing apart the functional neuroanatomy, even distinguishing it from the effects of pharmacologically active treatments.

See, for example, Brunstrom (2007) and Stockhorst, Enck, and Klosterhalfen

19 (2007). McKay and Dennett (2009) side with Humphrey (2002) in treating placebo “misbeliefs” as by-product, not adaptation. My speculations in this regard differ, but nothing critical to the central thesis turns on this difference.

20 Although if “medicine” is defined as “the provision of special care to the sick by others,” it might be very old (Evans, 2002: 459). But to establish my point it is only necessary that most medical claims concerning the cause of cures were groundless.

21 Some might contend that the apparent diminishing influence of institutional churches suggests diminished gullibility to beliefs that are not regulated in truth-conducive ways. Perhaps it is true that institutions of this sort are diminishing in influence—in some parts of the world—but that fact doesn’t imply a diminishing influence of such beliefs. The case of France might be instructive in this regard: it has experienced an ever dwindling supply of Roman Catholic clergy. According to tax authorities the number is now down to 36,000. But, according to those same tax authorities, France now has 40,000 professional astrologers (Kahane&Cavender, 2002: 137). If I am correct, a Conservation of Credulity principle seems to play an important role in human society.

22 In both the empirical and the philosophical literature, there is a general consensus that “self-deception” names an important phenomenon. But most discussion of this phenomenon is contentious. These controversies are fueled by the lack of convincing laboratory demonstrations (Paulhus, 2007) and by worries pertaining to specific interpretations, e.g. self-deception does not always help with the deception of others in the way that Trivers and other evolutionary psychologists have claimed (Van Leeuwen, 2007: 335). For purposes of this paper it is sufficient that self-deception illustrates both the generation of false beliefs and that it suggests one aspect of T-belief’s capacity for causally affecting psychological and social circumstance.

23 Reflection on some of the empirical evidence presented here has prompted others to wonder whether the relevant attitude is hope, not belief (Flanagan, 2002: 22, fn. 4). But hope seems ambiguous between belief and desire, not an amalgam (as will be described here). Hope in the sense of “hopeful” seems to be nothing more than belief in a better future (Breznitz, 1999: 629); and, most uses

24 of “I hope” seem synonymous with “I desire” or “I want.” Tertullian’s words are a common textbook example of irrationalism (Quine&Ullian, 1978: 60). 25 A balance must be maintained: too much positive illusion, and one will be disinclined to seek available help. Too little, and one might despair. Daniel Gilbert (2005: 177-178) refers to this as a “psychological immune system” that must, like its physiological counterpart, maintain a balance between hypo and

26 hyper activity. Other amalgams of belief and desire have been proposed, e.g. “besire” (Blackburn, 1998: 97-100).

27 Tamar SzabóGendler (2008) has recently introduced the concept “alief.” Like t-belief alief is claimed to “govern all sorts

of belief-discordant behavior” (2008: 663). But t-belief differs in several respects, including its ability to change the world to match its content.

28 Moore’s concern was to illustrate the distinction between what is asserted and what is implied; Wittgenstein bestowed the name “Moore’s paradox” (Baldwin,

29 1990: 226). Note that the same could not be said of unalloyed desire. “I want it to be the case that p” and “not-p” are not inconsistent (cf. Crane, 2001: 102- 105).

30 This is a testable implication of the t-belief hypothesis. 31 Richard Dawkins (1993) stakes out an especially uncompromising position: he

regards religious beliefs as either marks of cowardice or as “pernicious,” symptoms of disease such that those who hold them should be regarded as “patients.”

32 “External” and “internal” are best understood as occupying opposite ends of a 33 continuous spectrum.

A relevant example of this is the “tragedy of cognition” (Atran, 2003): we can

meta-represent self and others, project the future, and envision the demise of all 34 we care about. These too are part of the environment that we must adapt to.

I do claim that t-beliefs can be adaptive, in that they enhance fitness. Whether or not they count as biological adaptations (i.e. whether or not we have inherited them because they enhanced the fitness of our ancestors) is not something that needs to be dealt with here (cf. Buller, 2005: 35). But because false beliefs can seem so non-functional (cf. Konner, 2002: 15), and because maintaining a proper balance and calibration seems very complex (cf. Buller, 2005: 31-37), I suspect biological evolution may have played some role. Nevertheless, since no precision can as yet be given to the claims of usefulness here, it is better to allow that t-belief mechanisms are labile with the environment.

35 According to Owen Flanagan’s (1991: 32) Principle of Minimal Psychological Realism, our moral theories should not require of us that which is not possible

36 for creatures like us (also see Doris, 2002: 112). Note that placebo induction seems to be the only deliberate induction of a false belief that Wood (2008: 14) finds acceptable, albeit grudgingly: “To lie paternalistically to people may sometimes help them (for instance, to overcome a life-threatening illness), but . . . it shows a lack of respect . . . and seems 37 justifiable only temporarily, under very special conditions.”

“There are no matters about which we do not owe it both to ourselves and to others to maintain our intellectual integrity by forming our beliefs according to the evidence” (Wood, 2002: 33).

H. Results and Discussion, Part IV: Anti-Individualism and Vision Science.

I devoted five months to this project, including (a) exegesis of Tyler Burge’s recent work, (b) evaluation of relevant vision science materials, (c) drafting of the first written overview of the project, (d) correspondence with Tyler Burge, (e)

provision of funding from my NSC grant for 梁益堉's assistant, and (f) co-teaching a course at National Taiwan University, for which I neither received payment nor was accorded formal credit. This component of the three-year research project, in accord with prior agreements, is supposed to be written up by 梁益堉, since I wrote everything else that bears both of our names. For this component, I anticipate that, after it has been written up, I will serve as discussant, the role that he played on the already published manuscripts which bear his name, and that I will also be credited as co-author, sharing equal credit.

I. Results and Discussion, Part V: Partial and Whole Body Illusions.

i. Mental ownership constrains the rubber hand illusion.

(See Appendix 1)

ii. The malleability of self and body experiences.

(See Appendix 2)

iii. Self-specificity and mineness.

“...the decisive step in the making of consciousness is not the making of images and creating the basics of mind. The decisive step is *making the images ours*, making them belong to their right owners...” (Damasio 2010, p. 10)

Part I: Failure of the Self-Specificity Paradigm

1. A concern with mineness—“the respect in which mental states are experienced as *my own* states”—shared by analytic and continental philosophy, as well as by cognitive neuroscience.

2. Two experimental paradigms dedicated to research on mineness— both concerned to find that which *constitutes* the experiential self. 3. Self-Relatedness (SR)—focus on processing of stimuli “that are experienced as strongly related to one’s own person,” e.g. how we recognize some faces as our own, others as those of

famous people (Northoff et al. 2006). But SR seems not to adequately distinguish self from nonself.

4. Self-Specificity (SS):

Part I: Failure of the Self-Specificity Paradigm

A. Legrand and Ruby (2009) call for “paradigm shift.” Replace SR with SS.

B. Focus on what is most basic—self is distinct from nonself. “At the experiential level” self is “specific,” at least in the sense that “we can hardly help distinguishing between the self and everything else.”

C. Concentrate on “subjective perspective”—“the relating of a perceiving subject and a perceived object” (e.g. “my experience of biting a lemon”).

D. “Perspective is fundamentally a self-specifying process in the sense that it *constitutes* the self-nonsel self distinction.”

E. Concern with “being a self,” “minimal self,” “self-as-subject,” and “pre-reflective self.” No need for any explicit representation of self.

F. Operational definition of SS:

(i) Exclusivity: If a given self S is constituted by a SS component C, then C characterizes S exclusively. C could never characterize non-S.

(ii) Noncontingency: Loss of or change to C would result in loss of that distinction between S and non-S.

Subjective Perspective claimed “to meet both criteria”:

(i) Perspective is exclusive to self: two people can see the same thing, but neither perception can be reduced to the other, for they are had from different perspectives that differ systematically.

(ii) Perspective is noncontingent: any change, changes the self-nonsel self distinction.

G. Conclusion: “My perceptions, representations, and experiences are anchored in my perspective, and by virtue of this, they are mine rather than someone else’s or nobody’s.”

5. Counterexample to SS: “Double Visions”

A. Patient (DP) reports distress over “double visions” (Zahn et al. 2008). B. Turns out not to be “double vision”; instead, “he was able to see everything normally, but that he did not immediately recognize that he was the one who perceives and that he needed a second step to become aware that he himself was the one who perceives the object.” C. Symptoms restricted to visual object recognition. D. Apparent cause—hypometabolism in several areas, but “predominantly within right inferior

temporal and parieto-occipital regions.” E. When DP looks at a new object he satisfies both of SS’s operational conditions:

(i) Exclusivity: the image could not be constitutive of anyone who is not DP. (ii)

Noncontingency: change in that image would result in change to that particular distinction between DP and non-DP.

(iii) But from the satisfaction of these two conditions it does not necessarily follow that this is *DP’s* visual image. F. Similar reports from prodromal psychoses: e.g. patient who “reported that his feeling of his experiences as his own experiences only appeared a split-second delayed” (Sass and Parnas 2003, p. 438). G. Conclusion: Although the visual image is anchored in DP’s perspective, there is an important sense in which that perception or that experience is “nobody’s.” The same appears to be the case—pre-reflectively—for some cases of psychoses. Knowledge of the existence of a mental state is one thing; attribution of that mental state to a particular subject is something else.

Part II: The search for that which is uniquely constitutive of mineness is misguided; mineness is realized in multiple ways.

1. Previously shown that *access-distinction* (to a first approximation, introspection versus observation of the external world) does not account for mineness. Sometimes introspective access enables us to have a conscious experience only if we represent that experience as belonging to someone else (Lane and Liang 2011).

2. Above it has been shown that *subjective perspective* cannot secure mineness. 3. Varieties of Mineness: various phenomena show that mineness or its absence—ownership or disownership—can be realized in distinct ways.

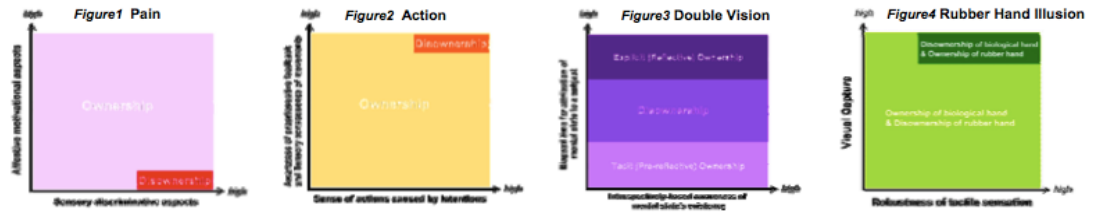
A. Subtraction of sensory experience: *disownership* in cortico-limbic disconnection syndrome (e.g. pain asymbolia) results when sensory-discriminative aspects retained while affective-motivational aspects eliminated. Pains “seem to belong to someone else, not to me” (Sierra 2009, p. 150). See Figure 1.

B. Addition of sensory experience: *disownership* of actions (e.g. passivity experiences) results when we are “abnormally aware” of proprioceptive feedback and the sensory consequences of movements (Frith 2005, p. 763). See Figure 2

C. Time delay: what seems to occur in the cases of “double visions” and prodromal psychoses is an abnormal time delay between (i) the introspectively-based knowledge of a mental state’s existence, and

(ii) the attribution of that mental state to a subject. See Figure 3. D. What happens in the case of the rubber hand illusion? See Figure 4.

4. Aberrant cases of mineness occur when tacit expectations are confounded. The four figures represent paired, but dissociable, mental states. In each case the confounding variable is represented by the vertical axis.



Conclusion:

1. A Negative Thesis: Failed attempts to identify that which is constitutive of mineness, leavened by familiarity with the varieties of mineness (including cases of aberrancy), suggest that there are no unique constituents.

2. A Positive Thesis: It may well be the case, however, that any attempt to construct an adequate explanatory framework of mineness will require inclusion of a Principle of Confounded Expectations (PCE).

iv. Mental ownership and the rubber hand illusion.

(See Appendix 3)

J. Results and Discussion, Part VI: The ethics of suicide research.

i. Media Impact on Individual Suicidality-A proposal for an ethical neuroimaging

study

(See Appendix 4)

K. Acknowledgments

To 戴華主任, of 國立成功大學人文社會科學中心, I owe my greatest debt, a debt that can never be adequately repaid. After I had all but abandoned hope of resurrecting an academic career, through his support and encouragement, I managed to summon the will to make one, last effort. Fortunately, this time, I was able to see a few projects through to their completion. Likewise, I am also deeply indebted to 林從一, 陳瑞麟, 王振寰, 鄭瑞城, and 關秉寅 and for their counsel and encouragement, aimed at getting me to restart my professional career, during 2006, at the conclusion of a lengthy series of unfortunate, personal events. As my professional career began to gain some measure of traction, 楊建銘, 吳建昌, and 葉素玲 also greatly aided my development through their wise counsel, keen insight, and willingness to collaborate with me, despite my many inadequacies. To 梁益琦, I express gratitude for his willingness to read and comment on portions of four manuscripts. I also thank him for having called an empirical example to my attention and for his assistance with the construction of a footnote. Finally, I will forever be grateful to Carl Hempel and to Julian Jaynes, who jointly encouraged me to pursue a professional career and a series of worthy projects, when I was in my early 20s. It was through no fault of theirs that results worthy of publication, meager though they are, have only recently seen their way into print. I hope that future publications can more adequately compensate them for the time, the guidance, and the inspiration that they so generously provided.

L. References

Acker, D., & Duck, N. (2008). Cross-cultural confidence and biased self-attribution. *The Journal of Social-Economics*, 37, 5: 1815-1824.

Agnew, J. H. W., & Webb, W. B. (1972). Measurement of sleep onset by EEG criteria. *American Journal of EEG Technology*, 12, 127-134.

Alloy, L. B. (1995). Depressive realism: Sadder but wiser? *The Harvard Mental Health Letter*, 11, 10: 4-5.

Alloy, L. B., & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives. In L. B. Alloy (Ed.), *Cognitive processes in depression* (pp. 223-265). New York: The Guilford Press.

Alloy, L. B., & Abramson, L. Y. (2007). Depressive realism. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 242-243). Los Angeles: Sage.

Alloy, L. B., & Ahrens, A. (1987). Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others. *Journal of Personality and Social Psychology*, 52, 2: 366-378.

Amrhein, C., & Schulz, H. (2000). Self reports after wakening – a contribution to sleep perception. *Somnologie*, 4(2), 61–67.

Anliker, J. (1966). Simultaneous changes in visual separation threshold and voltage of cortical alpha rhythm. *Science*, 153, 316–318.

Armor, D., & Taylor, S. (2003). The effects of mindset on behavior: Self-regulation in deliberative and implemental frames of mind. *Personality and Social Psychology Bulletin*, 29, 1: 86-95.

Arthur, W. (2004). *Biased embryos and evolution*. Cambridge, UK: Cambridge University Press.

Atran, S. (2003). The neuropsychology of religion. In R. Joseph (Ed.), *Neurotheology: Brain, science, spirituality, and religious experience* (pp. 147-166). Berkeley, CA: University Press.

Azekawa, T., Sei, H., & Morita, Y. (1990). Continuous alteration of EEG activity in human sleep onset. *Sleep Research*, 19, 7.

Baier, B. and H. Karnath. 2008. Tight link between our senses of limb ownership and self-awareness of actions. *Stroke: Journal of the American Heart Association* 39: 486–88.

Bakan, D. (1985). The apprehension of the placebo phenomenon. In L. White, B. Tursky, & G. E. Schwartz (Eds.), *Placebo: Theory, research, and mechanisms* (pp. 211-214). New York: The Guilford Press.

Baldwin, T. (1990). G. E. Moore. London: Routledge. Bausell, R. (2007). *Snake oil science: The truth about complementary and alternative medicine*. New York: Oxford University Press.

Balkin, T. J., Braun, A. R., Wesensten, N. J., Jeffries, K., Varga, M., Baldwin, P., et al (2002). The process of awakening: A PET study of regional brain activity

patterns mediating the re-establishment of alertness and consciousness. *Brain*, 125(10), 2308–2319.

Barrett, J., Lack, L., & Morris, M. (1993). The sleep-evoked decrease of body temperature. *Sleep*, 16(02), 93–99.

Baum, C. et al. 2000 Measuring function in Alzheimer's Disease. *Alzheimer's Quarterly*, Summer.

Baum, C. and D. Edwards 2003 What persons with Alzheimer's Disease can do. *Alzheimer's Quarterly* April/June.

Benedetti, F. (2008). Mechanisms of placebo and placebo-related effects across diseases and treatments. *The Annual Review of Pharmacology and Toxicology*, 48: 33-60.

Bennett, M. R., & Hacker, P. M. S. (2003). Philosophical foundations of neuroscience. Oxford: Blackwell. Beauregard, M. (2007). Mind does really matter: Evidence from neuroimaging studies of emotional self-regulation, psychotherapy, and placebo effect. *Progress in Neurobiology*, 81, 4: 218-236.

Black, I. B. 2001 *The Changing Brain: Alzheimer's Disease and advances in neuroscience*.

Blackburn, S. (1998). *Ruling passions*. New York: Oxford University Press.

Bonnet, M. H. (1986). Auditory thresholds during continuing sleep. *Biological Psychology*, 22(1), 3–10.

Bottini, G., E. Bisiach, R. Sterzi and G. Vallar. 2002. Feeling touches in someone else's hand. *NeuroReport* 13: 249–54.

Boyer, P. (2000). Evolution of the modern mind and the origins of culture: Religious concepts as a limiting case. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language, and meta-cognition* (pp. 93-112). New York: Cambridge University Press.

Braun, A. R., Balkin, T. J., Wesensten, N. J., Carson, R. E., Varga, M., Baldwin, P., et al (1997). Regional cerebral blood flow throughout the sleep–wake cycle. An H₂(15)O PET study. *Brain*, 120, 1173–1197.

Braun, A. R., Balkin, T. J., Wesensten, N. J., Gwady, F., Carson, R. E., Varga, M., et al (1998). Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. *Science*, 279(5347), 91–95.

Breen, N. et al. 2000. Towards an understanding of delusions of misidentification: four case studies. *Mind and Language* 15: 74–110.

- Breznitz, S. (1999). The effect of hope on pain tolerance. *Social Research*, 66, 2: 629-652.
- Brookfield, VT: Ashgate. Velleman, J. (2000). *The possibility of practical reason*. New York: Oxford University Press.
- Brunstrom, J. (2007). Associative learning and the control of human dietary behavior. *Appetite*, 49, 1: 268-271.
- Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge, MA: The MIT Press.
- Campbell, K., Bell, I., & Bastien, C. (1992). Evoked potential measures of information processing during natural sleep. In R. J. Broughton & R. D. Ogilvie (Eds.), *Sleep arousal and performance* (pp. 88–116). Boston, MA: Birkhauser.
- Case, T. (2000). *The need to believe: Motivation, disposition, and superstition*. Unpublished doctoral dissertation, Macquarie University, Australia.
- Case, T. I., Fitness, J., Cairns, D. R., Stevenson, R. J. (2004). Coping with uncertainty: superstitious strategies and secondary control. *Journal of Applied Social Psychology*, 34, 4: 848-871.
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.
- Church, A. T., Katigbak, M. S., del Prado, A. M., Valdez-Medina, J. L., Miramontes, L. G., & Ortiz, F. A. (2006). A cross-cultural study of trait self-enhancement, explanatory variables, and adjustment. *Journal of Research in Personality*, 40, 6: 1169-1201.
- Churchland, P. (in press). *Brain-based values*. New York: Oxford University Press.
- Clark, A. (1997). *Being there: Putting brain, body, and world together*. Cambridge, MA: The MIT Press.
- Clark, A. (2007). Curing cognitive hiccups: A defense of the extended mind. *The Journal of Philosophy*, 104, 4: 163- 192.
- Clark, A. 1993 *Associative Engines: Connectionism, Concepts, and Representational Change*. Cambridge: MIT Press.
- Clark, A. 1997 "The extended mind." *Analysis* 58(1), 7-19.
- Clark, A. 2002 "That special something: Dennett on the making of minds and selves." In Brook, A. and Ross, D., Eds. *Daniel Dennett*. New York: Cambridge University Press.
- Clark, A. 2003 *Natural-Born Cyborgs: minds, technologies, and the future of*

human intelligence. New York: Oxford Press. 2004 “Review Symposium: we have always been...cyborgs. Author’s Response.” *Metascience* 13: 169-181.

Clark, A. 2007a “Soft selves and ecological control.” In Spurrett, D. et al. eds. *Distributed Cognition and the Will*. MIT Press.

Clark, A. 2007b “Reinventing ourselves: the plasticity of embodiment, sensing, and mind.” *Journal of Medicine and Philosophy* 32: 263-282.

Clark, A. 2008 Pressing the flesh: a tension in the study of the embodied, embedded mind? *Philosophy and Phenomenological Research* 76(1): 37-59.

Clark, A. 2008 *Supersizing the Mind*. New York: Oxford University Press.

Colrain, I. M., Trinder, J., Fraser, G., & Wilson, G. V. (1987). Ventilation during sleep onset. *Journal of Applied Physiology*, 63, 2067–2074.

Colvin, C., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin*, 116, 1: 3-20.

Crane, T. (2001). *Elements of mind: An introduction to the philosophy of mind*. New York: Oxford University Press.

Cross, K. P. (1997). Not can, but will college teaching be improved? *New Directions for Higher Education*, 17: 1-15.

Czisch, M., Wetter, T. C., Kaufmann, C., Pollmacher, T., Holsboer, F., & Auer, D. P. (2002). Altered processing of acoustic stimuli during sleep: Reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *NeuroImage*, 16, 251–258.

Davidson, D. (1977). The method of truth in metaphysics. In P. A. Finch, T. E. Uehling, & H. K. Wettstein (Eds.), *Midwest studies in philosophy II: Contemporary perspectives in the philosophy of language*. Minneapolis, MN: University of Minnesota Press.

Davidson, D. (2003). Thought and talk. In T. O’Connor & D. Robb (Eds.), *Philosophy of mind: Contemporary readings* (pp. 355-369). London: Routledge.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., & Hobart, G. (1937). Changes in human brain potentials during the onset of sleep. *Science*, 86, 448–450.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., & Hobart, G. (1938). Human brain potentials during the onset of sleep. *Journal of Neurophysiology*, 1, 24–38.

Dawkins, R. (1993). Viruses of the mind. In B. Dahlbom (Ed.), *Dennett and his critics: Demystifying mind* (pp. 13-27). Oxford, MA: Blackwell.

De Gennaro, L., Ferrara, M., Ferlazzo, F., & Bertini, M. (2000). Slow eye movements and EEG power spectra during wake-sleep transition. *Clinical Neurophysiology*, 111(12), 2107–2115.

Diamond, J. (2001). *Snake oil and other preoccupations*. New York: Vintage.

Diederich, N., & Goetz, C. (2008). The placebo treatments in neurosciences: New insights from clinical and neuroimaging studies. *Neurology*, 71: 677-684.

Doris, J. (2002). *Lack of character*. Cambridge, UK: Cambridge University Press.

Evans, D. (2002). Pain, evolution, and the placebo response. *Behavioral and Brain Sciences*, 25, 4: 459-460.

Edelman, G. and G. Tononi 2000 *A Universe of Consciousness: how matter becomes imagination*. New York: Basic Books.

Evans, D. (2004). *Mind over matter in modern medicine*. London: HarperCollins.

Feinberg, T.E., A. Venneri, A.M. Simone, Y. Fan and G. Northoff. 2010. The neuro- anatomy of asomatognosia and somatoparaphrenia. *The Journal of Neurology, Neurosurgery & Psychiatry* 81: 276–81.

Finelli, L. A., Baumann, H., Borbély, A. A., & Achermann, P. (2000). Dual electroencephalogram markers of human sleep homeostasis: Correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience*, 101(3), 523–529.

Flanagan, O. (1991). *Varieties of moral personality: Ethics and psychological realism*. Cambridge, MA: Harvard University Press.

Flanagan, O. (2002). *The problem of the soul: Two visions of mind and how to reconcile them*. New York: Basic Books.

Flanagan, O. (2007). *The really hard problem: Meaning in a material world*. Cambridge, MA: The MIT Press.

Foulkes, D., & Vogel, G. (1965). Mental activity at sleep onset. *Journal of Abnormal Psychology*, 70, 231–243.

Frith, C. (1992). *The cognitive neurobiology of schizophrenia*. New Jersey: Lawrence Erlbaum Associates.

Galliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., et al. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92, 2: 325-336.

Gazzaniga, M. (2008). *Human: The science behind what makes us unique*. New York: HarperCollins.

Gendler, T. S. (2008). Alief and belief. *The Journal of Philosophy*, 105, 10: 634-663.

Gibson, E., Perry, F., Redington, D., & Kamiya, J. (1982). Discrimination of sleep onset stages: Behavioral responses and verbal reports. *Perceptual and Motor Skills*, 55(3 Pt 2), 1023–1037.

Gilbert, D. (1993). The assent of man: Mental representations and the control of belief. In D. Wegner & J. Pennebaker (Eds.), *Handbook of mental control* (pp. 57-87). Englewood- Cliffs, NJ: Prentice-Hall.

Gilbert, D. (2005). *Stumbling on happiness*. New York: Vintage Books.

Glannon, W. (2007). *Bioethics and the brain*. New York: Oxford University Press.

Goldberg, E. and D. Gougakov 2007 Goals, executive control, and action. In Baars, B. and N. Gage, Eds. *Cognition, Brain, and Consciousness*. New York: Academic Press.

Harsh, J., Voss, U., Hull, J., Schrepfer, S., & Badia, P. (1994). ERP and behavioral changes during the wake/sleep transition. *Psychophysiology*, 31(3), 244–252.

Heine, S. (2001). Self as cultural product: An examination of East Asian and North American selves. *Journal of Personality*, 69, 6: 881-906.

Held, V. (1996). Whose agenda? Ethics versus cognitive science. In L. May, M. Friedman, & A. Clark (Eds.), *Mind and morals: Essays on ethics and cognitive science*. Cambridge, MA: The MIT Press.

Hori, T. (1985). Spatiotemporal changes of EEG activity during waking–sleeping transition period. *International Journal of Neuroscience*, 27(1), 101–114.

Hori, T., Hayashi, M., & Hibino, K. (1992). An EEG study of the hypnagogic hallucinatory experience. *International Journal of Psychology*, 27, 420.

Hori, T., Hayashi, M., & Morikawa, T. (1991). Changes of EEG patterns and reaction time during hypnagogic state. *Sleep Research*, 20, 20.

Hori, T., Hayashi, M., & Morikawa, T. (1994). Topographical EEG changes and the hypnagogic experience. In R. D. Ogilvie & J. R. Harsh (Eds.), *Sleep onset*(pp. 237–253). Washington, DC: American Psychological Association.

Humphrey, N. (2002). *The mind made flesh: Essays from the frontiers of psychology and evolution*. New York: Oxford University Press.

Humphrey, N. (2004). Placebo effect. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (pp. 735-736). New York: Oxford University Press.

Hurlburt, R., & Schwitzgebel, E. (2007). *Describing inner experience: Proponent meets skeptic*. Cambridge, MA: The MIT Press.

Jacobson, A., Kales, A., Lehmann, D., & Hoedemaker, F. S. (1964). Muscle tonus in human subjects during sleep and dreaming. *Experimental Neurology*, 10(5),418–424. Johnson, L. C. (1970). A psychophysiology for all states. *Psychophysiology*, 6, 501–516.

Joyce, R. (2007). *The evolution of morality*. Cambridge, MA: The MIT Press.

Juarro, A. 2004 Commentary for Review Symposium on “We have always been...cyborgs.” *Metascience* 13: 149-153.

Kahan, T., LaBerge, S., Levitan, L., & Zimbardo, P. (1997). Similarities and differences between dreaming and waking cognition: An exploratory study. *Consciousness and Cognition*, 6, 132–147.

Kahane, H., & Cavender, N. (2002). *Logic and contemporary rhetoric* (9th ed.). Belmont, CA: Wadsworth.

Kahn, D., & Hobson, J. A. (2005). State-dependent thinking: A comparison of waking and dreaming thought. *Consciousness and Cognition*, 14, 429–438.

Kaufmann, C., Wehrle, R., Wetter, T. C., Holsboer, F., Auer, D. P., Pollmächer, T., et al (2006). Brain activation and hypothalamic functional connectivity during human non-rapid eye movement sleep: An EEG/fMRI study. *Brain*, 129(3), 655–667.

Keinan, G. (2002). The effects of stress and desire for control on superstitious behavior. *Personality and Social Psychology Bulletin*, 28, 1: 102-108.

Kennett, J., & Fine, C. (2008). Internalism and the evidence from psychopaths and “acquired sociopaths.” In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development* (pp. 173-190). Cambridge, MA: The MIT Press.

Kjaer, T. W., Nowak, M., & Lou, H. C. (2002). Reflective self-awareness and conscious states: PET evidence for a common midline parietofrontal core. *NeuroImage*, 17(2), 1080–1086.

Konner, M. (2002). *The tangled wing: Biological constraints on the human spirit* (2nd ed.). New York: Henry Holt.

Krauchi, K., Cajochen, C., Werth, E., & Wirz-Justice, A. (2000). Functional link between distal vasodilation and sleep-onset latency. *The American Journal of Physiology – Regulatory, Integrative and Comparative Physiology*, 278(3), R741–748.

Kriegel, U. 2005. Naturalizing subjective character. *Philosophy and Phenomenological Research* Vol. LXXI, No. 1: 23–57.

Lane, T. and C. Liang. 2008. Higher-order thought and the problem of radical confabulation. *The Southern Journal of Philosophy* 46: 69–98.

Lehmann, D., Grass, P., & Meier, B. (1995). Spontaneous conscious covert cognition states and brain electric spectral states in canonical correlations. *International Journal of Psychophysiology*, 19, 41–52.

Levy, N. (2007). *Neuroethics*. New York: Cambridge University Press.

Lex, B. W. (1977). Voodoo death: New thoughts on an old explanation. In D. Landy (Ed.), *Culture, disease, and healing studies in medical anthropology* (pp. 327-331). New York: Macmillan.

Liang, C. and T. Lane. 2009. Higher-order thought and pathological self: the case of somatoparaphrenia. *Analysis* 69: 661–68.

Litchman, J. (1974). *Mentals EMG in human sleep and wakefulness*. Chicago: University of Chicago.

Lynch, M. (2004). *True to life: Why truth matters*. Cambridge, MA: The MIT Press.

Maquet, P. (2000). Functional neuroimaging of normal human sleep by positron emission tomography. *Journal of Sleep Research*, 9(3), 207–231.

Matsuzawa, T. 2007 Comparative cognitive development. *Developmental Science* 10(1): 97-103.

Mayberg, H. S., Silva, J. A., Brannan, S. K., TeKell, J. L., Mahurin, R. K., & McGinnis, S., et al. (2002). The functional neuroanatomy of the placebo effect. *The American Journal of Psychiatry*, 159, 5: 728-737.

McGinn, C. (2004). *Mindsight: Image, dream, meaning*. Cambridge, MA: Harvard University Press.

McKay, R., & Dennett, D. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32, 6: 493-561.

McKendree-Smith, N., & Scogin, F. (2000). Depressive realism: Effects of depression severity and interpretation time. *Journal of Clinical Psychology*, 56, 12: 1601-1608.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20: 91-136.

Mele, A. R. (2001). *Self-deception unmasked*. Princeton, NJ: Princeton University Press.

Melzack, R. 1989. Phantom limbs, the self, and the brain. *Canadian Psychology* 30: 1–16.

Merica, H., Fortune, R. D., & Gaillard, J. M. (1991). Hemispheric temporal organization during the onset of sleep in normal subjects. In M. G. Terzano, P. L. Halasz, & A. C. Declerck (Eds.), *Phasic events and dynamic organization of sleep*. New York: Raven Press.

Mitchell, L.A., R.A. MacDonald and E.E. Brodie. 2004. Temperature and the cold pressor test. *The Journal of Pain* 5: 233–37.

Muller, H. J. 1997 (original 1950) “The penalty for relaxing natural selection.” In Ridley, M., ed. *Evolution*. New York: Oxford University Press.

Muraven, M., & Baumeister, R. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126, 2: 247-259.

Muzur, A., Pace-Schott, E. F., & Hobson, J. A. (2002). The prefrontal cortex in sleep. *Trends in Cognitive Sciences*, 6, 475–481.

Naifeh, K. H., & Kamiya, J. (1981). The nature of respiratory changes associated with sleep onset. *Sleep*, 4(1), 49–59.

Niemi, M. (2009). Cure in the mind. *Scientific American Mind*, 20, 1: 42-49.

Nofzinger, E. A., Buysse, D. J., Miewald, J. M., Meltzer, C. C., Price, J. C., Sembrat, R. C., et al (2002). Human regional cerebral glucose metabolism during non-rapid eye movement sleep in relation to waking. *Brain*, 125(5), 1105–1115.

Nofzinger, E. A., Mintun, M. A., Wiseman, M., Kupfer, D. J., & Moore, R. Y. (1997). Forebrain activation in REM sleep: An FDG PET study. *Brain Research*, 770(1–2), 192–201.

Nozick, R. (1993). *The nature of rationality*. Princeton, NJ: Princeton University Press.

Nozick, R. (1989). *The examined life: Philosophical meditations*. New York: Simon & Schuster.

Ogilvie, R. D., & Simons, I. (1992). Falling asleep and waking up: A comparison of EEG spectra. In R. J. Broughton & R. D. Ogilvie (Eds.), *Sleep, arousal and performance* (pp. 73–87). Boston: Birkhäuser.

Ogilvie, R. D., Simons, I. A., Kuderian, R. H., MacDonald, T., & Rustenburg, J. (1991). Behavioral, event-related potential, and EEG/FFT changes at sleep onset. *Psychophysiology*, 28(1), 54–64.

- Ogilvie, R. D., & Wilkinson, R. T. (1984). The detection of sleep onset: Behavioral and physiological convergence. *Psychophysiology*, 21(5), 510–520.
- Ogilvie, R. D., & Wilkinson, R. T. (1988). Behavioral versus EEG-based monitoring of all-night sleep/wake patterns. *Sleep*, 11(02), 139–155.
- Ogilvie, R. D., Wilkinson, R. T., & Allison, S. (1989). The detection of sleep onset: Behavioral, physiological, and subjective convergence. *Sleep*, 12(5), 458–474.
- Oken, B. (2008). Placebo effects: Clinical aspects and neurology. *Brain*, 131, 11: 2812-2823.
- Paulhus, D. L. (2007). Self-deception. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 189-190). Los Angeles: Sage.
- Peigneux, P., Salmon, E., Garraux, G., Laureys, S., Willems, S., & Dujardin, K. (2001). Neural and cognitive bases of upper limb apraxia in corticobasal degeneration. *Neurology*, 57(7), 1259–1268.
- period? *Consciousness and Cognition*, 19, 1084–1092.
- Porte, H. S. (2004). Slow horizontal eye movement at human sleep onset. *Journal of Sleep Research*, 13, 239–249.
- Postal, K.S. 2005. The mirror sign delusional misidentification symptom. In *The Lost Self: Pathologies of the Brain and Identity*, eds T. Feinberg and J. Keenan, 131–46. New York: Oxford University Press.
- Poundstone, W. 1992 *Prisoner's Dilemma*. New York: Doubleday.
- Prinz, J. J. (2007). *The emotional construction of morals*. New York: Oxford University Press.
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *TRENDS in Cognitive Science*, 11, 1: 37-43.
- Pronin, E. (2008). How we see ourselves and how we see others. *Science*, 320, 5880: 1177-1180.
- Pronin, E., & Kugler, M. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43: 565-578.
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28, 3: 369-381.
- Quine, W. V., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). New York: McGraw-Hill.
- Quine, W. V. O. (1994). Natural kinds. In H. Kornblith (Ed.), *Naturalizing epistemology* (pp. 57-76). Cambridge, MA: MIT Press.

Railton, P. (2003). *Facts, values, and norms: Essays toward a morality of consequence*. Cambridge, UK: Cambridge University Press.

Ramsøy, T. 2007 Methods for observing the living brain. In Baars, B. and N. Gage, Eds. *Cognition, Brain, and Consciousness*. New York: Academic Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Rechtschaffen, A. (1994). Sleep onset: Conceptual issues. In R. D. Ogilvie & J. R. Harsh (Eds.), *Sleep onset* (pp. 3–17). Washington, DC: American Psychological Association. Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology, techniques and scoring for sleep stages of human subjects*. Los Angeles: University of California.

Ridderinkhof, K. et al. 2002 Perservative behavior in older adults: performance-monitoring, rule-induction, and set shifting. *Brain Cognition* 49(3): 382-401.

Rosenthal, D.M. 2002. Explaining consciousness. In *Philosophy of Mind*, ed. D.J. Chalmers, 406–21. New York: Oxford University Press.

Rosenthal, D.M. 2010. Consciousness, the self and bodily location. *Analysis* 70: 270–76. Sierra, M. and G. Berrios. 2000. The Cambridge Depersonalization Scale: a new instrument for the measurement of depersonalization. *Psychiatry Research* 93: 153–64. Wittgenstein, L. 1969. *The Blue and the Brown Books*. Oxford: Blackwell.

Rosenthal, D.M. 1997. A theory of consciousness. In *The Nature of Consciousness*, eds N. Block, O. Flanagan, and G. Guzeldere, 729–53. Cambridge, MA: MIT Press.

Rosenthal, D.M. 2002a. Explaining consciousness. In *Philosophy of Mind*, ed. D. J. Chalmers, 406–421. New York: Oxford University Press.

Rosenthal, D.M. 2002b. Persons, minds, and consciousness. In *The Philosophy of Marjorie Grene*, eds R.E. Auxier and L.E. Hahn, 199–220. Chicago, Illinois: Open Court.

Rosenthal, D.M. 2004. Being conscious of ourselves. *The Monist* 87: 161–84.

Rosenthal, D.M. 2005. *Consciousness and Mind*. New York: Oxford University Press.

Rowley, J. T., Stickgold, R., & Hobson, J. A. (1998). Eyelid movements and mental activity at sleep onset. *Consciousness and Cognition*, 7(1), 67–84.

Scanlon, T. J., Luben, R. N., Scanlon, F. L., & Singleton, N. (1993). Is Friday the 13th bad for your health? *British Medical Journal*, 307, 6919: 1584-1586.

Schacter, D. L. (1976). The hypnagogic state: A critical review of the literature. *Psychological Bulletin*, 83(3), 452–481.

Scott, D. J., Stohler, C. S., Egnatuk, C. M., Wang, H., Koeppe, R. A., & Zubieta, J. K. (2007). Individual differences in reward responding explain placebo-induced expectations and effects. *Neuron*, 55, 2: 325-336.

Searle, J. (1992). *The rediscovery of mind*. Cambridge, MA: The MIT Press.

Searle, J. (2001). *Rationality in action*. Cambridge, MA: The MIT Press.

Sewitch, D. E. (1984). The perceptual uncertainty of having slept: The inability to discriminate electroencephalographic sleep from wakefulness. *Psychophysiology*, 21(3), 243–259.

Shah, N., & Velleman, J. (2005). Doxastic deliberation. *The Philosophical Review*, 114, 4: 497-534.

Shermer, M. (2002). *Why people believe weird things: Pseudoscience, superstition, and other confusions of our time* (Rev. ed.). New York: Henry Holt.

Shoemaker, S. 1968. Self-reference and self-awareness. *The Journal of Philosophy* 65: 555–67.

Simon, C. W., & Emmons, W. H. (1956). Responses to material presented during various levels of sleep. *Journal of Experimental Psychology*, 51, 89–97.

Sinnott-Armstrong, W. (Ed.). (2008). *Moral psychology*, Vols. 1-3. Cambridge, MA: The MIT Press.

Sobel, D., & Copp, D. (2001). Against direction of fit accounts of belief and desire. *Analysis*, 61, 1: 44-53.

Stanovich, K. E. 2004 *The robot's rebellion: finding meaning in the age of Darwin*. Chicago: The University of Chicago Press.

Sterelny, K. 2004 Externalism, epistemic artifacts, and the extended mind. In Schantz, R. Ed. *The Externalist Challenge: New Studies on Cognition and Intentionality*. New York: de Gruyter. Vellman, J. 2000 *The Possibility of Practical Reason*. New York: Oxford University Press.

Stockhorst, U., Enck, P., & Klosterhalfen, S. (2007). Role of classical conditioning in learning gastrointestinal symptoms. *World Journal of Gastroenterology*, 13, 25: 3430-3437.

Suhler, C., & Churchland, P. S. (2009). Control: Conscious and otherwise. *Trends in Cognitive Sciences*, 13, 341–347.

Talmon-Kaminski, K. (2008). In a mirror, darkly: Does superstition reflect rationality? *Skeptical Inquirer*, 32, 4: 48-51.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 2: 193-210.

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin*, 116, 1: 21-27.

Taylor, S. E., & Gollwitzer, P. M. (1995). The effects of mindset on positive illusions. *Journal of Personality and Social Psychology*, 69, 2: 213-226.

Taylor, S. E., Kemeny, M. E., Reed, G. M., Bower, J. E., & Gruenewald, T. L. (2000). Psychological resources, positive illusions, and health. *American Psychologist*, 55, 1: 99-109.

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Are self-enhancing cognitions associated with healthy or unhealthy biological profiles? *Journal of Personality and Social Psychology*, 85, 4: 605-615.

Tertullianus, Quintus Septimius Florens (2010). Tertullianproject: "De Carne Christi," chapter 5, verse 4. Retrieved August 7, 2010, from <http://www.tertullian.org/>

Tiger, L. (1985). *Optimism: The biology of hope* (2nd ed.). New York: Kodansha.

Tiger, L. (1999). Hope springs internal. *Social Research*, 66, 2: 611-623.

Trivers, R. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, 907: 114-131.

Tirgu-Mures, Rumania, and Basel, Switzerland: S. Kager. Wehr, T. A. (1990). Effects of wakefulness and sleep on depression and mania. In J. Mounplaisir & R. Godbout (Eds.), *Sleep and biological rhythms: Basic mechanisms and applications to psychiatry* (pp. 42-86). Oxford: Oxford University Press.

Tsuno, N., Shigeta, M., Hyoki, K., Kinoshita, T., Ushijima, S., Faber, P. L., et al (2002). Spatial organization of EEG activity from alertness to sleep stage 2 in old and younger subjects. *Journal of Sleep Research*, 11(1), 43-51.

Vallar, G. and R. Ronchi. 2009. Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research* 192: 533-51.

Van Leeuwen, D. (2007). The spandrels of self-deception: Prospects for a biological theory of a mental phenomenon. *Philosophical Psychology*, 20, 3: 329-348.

Velleman, J. (1999). The possibility of practical reason. In R. Jay Wallace (Ed.), *Reason, emotion, and will* (pp. 185-218).

Vogel, G., Barrowclough, B., & Giesler, D. (1972). Limited discriminability of REM and sleep onset reports and its psychiatric implications. *Archives of General Psychiatry*, 26, 449–455.

Vogel, G., Foulkes, D., & Trosman, H. (1966). Ego functions and dreaming during sleep onset. *Archives of General Psychiatry*, 14, 238–248.

Vyse, S. (2000). *Believing in magic: The psychology of superstition*. New York: Oxford University Press.

Wackermann, J. (2006). Rationality, universality, and individuality in a functional conception of theory. *International Journal of Psychophysiology*, 62, 411–426.

Wackermann, J., Pütz, P., Büchi, S., Strauch, I., & Lehmann, D. (2002). Brain electrical activity and subjective experience during altered states of consciousness: Ganzfeld and hypnagogic states. *International Journal of Psychophysiology*, 46, 123–146.

Webb, W. B. (1980). The natural onset of sleep. In B. A. L. Popoviciu & G. Badia (Eds.), *Sleep 1978. Fourth European congress on sleep research* (pp. 19–23).

Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 16: 267-297.

Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: The MIT Press.

Weigand, D., Michael, L., & Schulz, H. (2007). When sleep is perceived as wakefulness: An experimental study on state perception during physiological sleep. *Journal of Sleep Research*, 16(4), 346–353.

Wenger, A., & Fowers, B. (2008). Positive illusions in parenting: Every child is above average. *Journal of Applied Social Psychology*, 38, 3: 611-634.

Wilkes, K.V. 1993. *Real People: Personal Identity without Thought Experiments*, New York: Oxford University Press.

Williams, B. (1973). *Problems of the self*. Cambridge, UK: Cambridge University Press.

Williams, B. (2002). *Truth and truthfulness: An essay in genealogy*. Princeton, NJ: Princeton University Press.

Wood, A. (2002). *Unsettling obligations: Essays on reason, reality, and the ethics of belief*. Stanford, CA: Center for the Study of Language and Information.

Wood, A. (2008). The duty to believe according to the evidence. In E. T. Long & P. Horn (Eds.), *Ethics of belief: Essays in tribute to D. Z. Phillips* (pp. 7-24). Dordrecht, the Netherlands: Springer.

Yang, C.-M., Han, H. Y., Yang, M. H., Su, W. C., & Lane, T. (2010). What subjective experiences determine the perception of falling asleep during sleep onset

Zubieta, J., & Stohler, C. (2009). Neurobiological mechanisms of placebo responses. *The New York Academy of Science*, 1156: 198-210.

5. Self-evaluation of project outcome

國科會補助專題研究計畫成果報告自評表

<p>請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。</p>
<p>1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 X 達成目標</p> <p><input type="checkbox"/> 未達成目標（請說明，以 100 字為限）</p> <p><input type="checkbox"/> 實驗失敗</p> <p><input type="checkbox"/> 因故實驗中斷</p> <p><input type="checkbox"/> 其他原因</p> <p>說明：</p>
<p>2. 研究成果在學術期刊發表或申請專利等情形：</p> <p>論文：X 已發表 <input type="checkbox"/> 未發表之文稿 <input type="checkbox"/> 撰寫中 <input type="checkbox"/> 無</p> <p>專利：<input type="checkbox"/> 已獲得 <input type="checkbox"/> 申請中 <input type="checkbox"/> 無</p> <p>技轉：<input type="checkbox"/> 已技轉 <input type="checkbox"/> 洽談中 <input type="checkbox"/> 無</p> <p>其他：（以 100 字為限）</p>
<p>請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限） Here I indicate only the principal, already published, contribution:</p>

在心理學、神經科學尚未發展的六〇年代，哲學家要研究心理或心靈，僅能透過內省或純思考方式。美國哲學家 Sydney Shoemaker 在 1968 年提出「IEM (immunity to error through misidentification)」學說，他認為當「我」做為主體，只要透過內省來感知痛覺、觸覺等感官經驗，那必然是「我」的感受，不可能辨識錯誤。例如當一個人說自己牙痛的時候，我們並不會質疑對方「那真的是你的痛覺嗎？」此外，Shoemaker 主張此關係不僅維持我們的體感經驗，同時也作用於行動意識與視覺感知。

然而，本研究透過實際病例與實驗結果，證明此關係並非必然。以研究病患症狀為例，某些病患會將自己的肢體視為「他人的」，而產生主體與意識經驗的分離。舉例而言，某病患將自己的左手視為外甥女，實驗設計反覆觸碰其左手，當病患被告知其左手要被觸碰時，她表示並無感覺，只有當實驗者告知她「外甥女的左手要被觸碰了」，病患才有觸覺反應。此病例說明雖然「我」是主體，卻必須將意識經驗表徵為「他人的」，才能透過內省去體驗並恢復感知。

本研究以此成功推翻了長年以來廣為接受的 Shoemaker 學說 (IEM)，認為其學說僅能視為「假說」，並非任何情況下都能成立。本研究旨在以心理學、神經科學等研究方法來檢驗哲學問題，並期望達成以下目標：(1) 透過告知臨床醫生哪些問題應被問及，協助醫生更了解病患情況；(2) 設計實驗使我們更了解「自我」、「意識經驗」與「身體」之間的關係。

6. Appendix

ASSC 14 Toronto

Mental Ownership Constrains the Rubber Hand Illusion

Timothy Lane
National Chengchi University
June 26, 2010

1



Our talk today

- 1. Two dubious assumptions and a working hypothesis
- 2. Pilot study in Taiwan
- 3. Philosophical and empirical implications
- 4. Neural mechanisms
- 5. Conclusions

3

Part I. Dubious assumptions about self-consciousness

- First, body ownership is uniquely determined by introspection.
- Second, the Wittgenstein-Shoemaker (1968) question is absurd—"there is no question of recognizing a person when I say I have tooth-ache. To ask 'are you sure it is *you* who have pains?' would be nonsensical."

4

First Dubious Assumption

- E.G. Brewer (1995): "...how we experience our body *as ours*. Clearly this is not, and cannot be, an external perceptual phenomenon. For **when we perceive it from the outside, our body has no indelible stamp of ownership**. It appears just as one object among many..."

5

Problem with First Dubious Assumption

- Although the "indelible stamp of ownership" is more likely associated with introspection than with external perception, **this is contingent**.
- What we introspect can feel alien, or fail to be represented as belonging to self: e.g. depersonalization in florid schizophrenic episodes.
- When we perceive others experiencing mental states, those states can be represented as belonging to self: e.g. vision-touch synesthesia.

6

Second Dubious Assumption

- Our claim: The Wittgenstein-Shoemaker Question should be asked—‘are you sure it is *you* who is having this experience?’
- Why?
- **Somatoparaphrenia** (Bottini, 2002).
- **Body Swap**: “I was shaking hands with myself.” (Petkova & Ehrsson, 2008)
- **Atypical “double-vision”** (Zahn, et al, 2009)

But, hold onto a working hypothesis

- The distinction between **Self-as-subject** and **Self-as-object** helps explain aspects of self-consciousness.
- Self-as-subject: I am the one who is in pain.
- Self-as-object: I am the one who is bleeding.

Part II. Phenomenology of RHI

- Conflicting tactile and visual stimuli lead to proprioceptive drift and touch referral.
- And many subjects feel that the rubber hand is their hand.
- Here we focus on touch referral—“I feel tactile sensations on the rubber hand.”

Phenomenology of RHI

- Psychometric approaches, e.g. PCA, to introspective reports (Longo et al. 2008).
- Attempt to evoke and quantify structures of experience.

Tsakiris' Empirical Model (2010)

```

    graph TD
      BM[Body-model] --> BS[Body-state]
      BS --> T((Touch))
      BS --> VT((Vision of Touch))
      T --> TR((Touch Referral))
      VT --> P((Posture))
      P --> VF((Visual Form))
      TR --> BO[Body-ownership]
      VF --> BO
      T --> VT
      VT --> P
      P --> VF
      VF --> BO
  
```

Two concerns

- **Concern 1** Longo et al. ask 27 questions, but neglect the distinction between **self-as-subject** and **self-as-object**. Their questions concern only **self-as-object**.
- **Concern 2** **Body ownership** concerns **self-as-object**, whether a body part (e.g. a hand) or a full body belongs to me. **Mental ownership** concerns **self-as-subject**, whether I represent myself as the unique subject of experiences. **Body ownership** and **mental ownership** are distinct.

Two suggestions

- Longo et al. asked: "it seemed like the touch I felt was caused by the paintbrush touching the rubber hand." (2008, 983)
- Longo's question is actually about the cause, not about where the sensation is felt.
- **Suggestion 1:** A subjective touch referral question should be asked: "I felt the touch on the rubber hand." Responses then judged on a Likert Scale: from +3 ("strongly agreed") to -3 ("strongly disagreed").

13

Our suggestions

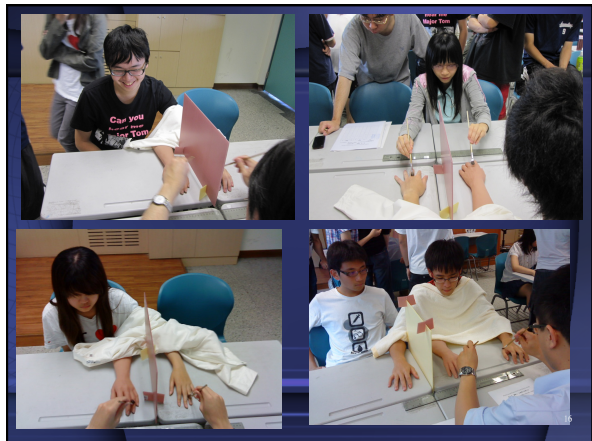
- **Suggestion 2:** Questions pertaining to **mental ownership** and its relation to body ownership should be asked.
- For example, "it seemed like I was the one who was feeling brush strokes on the rubber hand."
 - ♦ "I felt that the rubber hand was mine before I felt that the touch on the rubber hand was mine."
 - ♦ "I felt that the touch on the rubber hand was mine before I felt that the rubber hand was mine."

14

Our Predictions

1. Only after one approaches "strongly agree" (+2 or +3) on **subjective touch referral** will rubber hand ownership be experienced.
2. **Mental ownership**, ceteris paribus, is a prerequisite for **body ownership**.

15



Pilot study result 1 (only +2 or +3):

- 20 out of 66 subjects reported **body ownership** of rubber hand.
- 29 out of 66 subjects reported experiencing **touch referral**.
- 22 out of 66 subjects reported **mental ownership** of tactile sensations on the rubber hand.

17

Pilot study result 1:

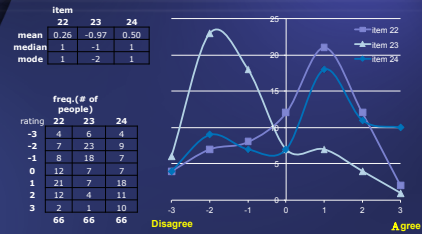
- As predicted, **body ownership** only obtains when experiencing strong (+2 or +3) sensation of **touch referral**. (Only four anomalous cases.)
- As predicted, **mental ownership** correlates with **body ownership**. (Only one anomalous case.)

18

Pilot study result 2:

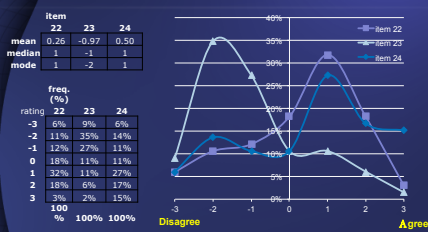
- As we predicted, **mental ownership precedes body ownership**.
- It seems to be the case that **mental ownership is a constraint on body ownership**.

- 22. 在實驗過程之中，我覺得自己像個旁觀者在觀察自己的感受
- 23. 我先感覺到那橡膠手臂是我的，才覺得那橡膠手臂被刷的觸覺是我的
- 24. 我先感覺到橡膠手臂被刷的觸覺是我的，才覺得那橡膠手臂是我的



X axis: ratings, Y axis: # of people having that rating.

- 22. 在實驗過程之中，我覺得自己像個旁觀者在觀察自己的感受
- 23. 我先感覺到那橡膠手臂是我的，才覺得那橡膠手臂被刷的觸覺是我的
- 24. 我先感覺到橡膠手臂被刷的觸覺是我的，才覺得那橡膠手臂是我的



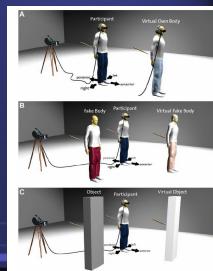
X axis: ratings, Y axis: % of people having that rating.

Part III. Philosophical and Empirical Implications

- The Self-Model Theory of Subjectivity (SMT) by Metzinger (2003, 2008, 2009).
- “Mineness”: “a higher order property of particular forms of phenomenal content ... an immediately given, non-conceptual sense of ownership.”
- Examples: “Subjectively, my leg is always experienced as being a part of me”; “My thoughts and feelings are always experienced as part of my own consciousness”; “My volitional acts are always initiated by myself.”

The Self-Model of Subjectivity (SMT) applied to the full-body illusion

- Metzinger regards self-identification as identification of a virtual body as one’s own body. **Self-identification** is taken to be “a globalized form of **identification with the body as a whole** (as opposed to ownership for body parts)” Metzinger and Blanke (2008, p. 8).



Our concerns

- Like Longo et al. (2008), Metzinger does not clearly distinguish self-as-subject from self-as-object.
- Our pilot study suggests that the distinction should be made.
- To understand self-consciousness there is a need to consider self-as-subject.

An empirical issue

- **Distinct Mechanism** (Metzinger and Blanke 2008, p. 9): “neural mechanisms of partial ownership and self-identification are...likely to differ.”
- **Shared Mechanism** (Tsakiris 2009, p. 9): “...the necessary conditions for the experience of ownership over a body-part seem to be the same as the ones involved in the experience of ownership over full bodies... available empirical findings from the two domains suggest that very similar neurocognitive process are involved in ownership of body-parts and bodies.”

25

Our current position

- On this issue, we side with the Shared Mechanism Thesis.
- Why? From the perspective of mental ownership the distinction between body-part and full-body is not so important.
- What matters is that the sense of who is the subject of synchronous sensations remains constant.

26

Our current position

- Note that even for full-body illusions, only a small portion of the back is being stroked. What matters, what is critical to “self-identification” (where self is understood as subject, not object), is that the subject of synchronous stroking remains constant.
- Burden of proof resides with advocates of Distinct Mechanism Thesis.

27

Part IV. Mechanism of Self-as-Subject—Initial Thoughts

- Most discussions seem only concerned with self-as-object. Adequate explanation of these phenomena will require investigation of the self-as-subject mechanism.
- **Candidate 1.** Northoff and Bermpohl (2006): the neural substrate in virtue of which self “mediates ownership of experience” are the cortical midline structures (CMS), that lie within the default mode network. (BA: 7, 9, 10, 11, 12, 23, 24, 25, 26, 29, 30, 31, and 32.)

28

Link between CMS and self-as-subject remains weak

- More relevant to the narrative self—my personality and my relatedness to others.
- Perhaps too inclusive—reading of other minds, memory recall, inductive and deductive reasoning, resting state, etc. And emphasis on mental content rather than cognitive processing (Legrand and Ruby 2009).

29

Mechanism of Self-as-Subject

- **Candidate 2.** A right fronto-parietal network (Baars et al. 2003), that overlaps with areas containing mirror neurons. (Inferior frontal cortex and the rostral portion of the inferior parietal lobule.)
- More generally: non-conscious sensory processing becomes conscious when a fronto-parietal network is engaged.

30

Mechanism of Self-as-Subject

- Conscious awareness of stimulus involves frontward spread of activation beyond sensory regions of the posterior cerebrum.
- In describing this “spread of activation” we think, rather cautiously, in terms of functional connectivity. Possibly realized in virtue of synchronized oscillatory neuronal responses.

31

The Mechanism—initial thoughts

- **With Regard to RHI.** Posterior parietal cortex (PPC) “is involved in the resolution of the conflict between the incoming visual and tactile information, and the resulting recalibration of the visual and tactile coordinate systems.” (Tsakiris 2010, p. 7)
- Provisionally, visual capture is realized in virtue of functional connectivity between the right frontal-parietal network and the PPC.

32

Conclusions

1. The Wittgenstein-Shoemaker Question should be asked.
2. The distinction between Self-as-subject and Self-as-object should be retained as a working hypothesis.
3. Pilot study supports **mental ownership as a constraint on body ownership.**
4. Burden of proof falls to advocates of Distinct Mechanism Thesis.
5. There is a need to investigate the mechanism that underlies self-as-subject.

33

The Malleability of Self and Body Experience

October 29, 2010

Su-Ling Yeh,
Center for Neurobiology and Cognitive Science & Department
of Psychology, National Taiwan University
Timothy Lane,
Research Center for Mind, Brain, and Learning, National
Chengchi University

Outline

- Naturally Occurring Instances of Malleability.
- Recently Discovered Experimental Paradigms.
- Current Consensus
- Mental Ownership Hypothesis.
- Our Experimental Probes.
- Practical Applications.
- Conclusion

NatO: Phantom Sensations

- Described in medical literature more than 500 years ago.
- Commonplace in amputees.
- Even in congenital cases (10-20%).
- Still neglected by medical practitioners.
- Although it is well-known, remains highly counter-intuitive.

NatO: Out of Body Experience

- Representation of one's body from an impossible, third-person perspective (e.g. seeing oneself from above)
- Occurrence: sleep paralysis, surgery, during severe accidents, seizures due to lesions or dysfunctions in the temporoparietal junction (TPJ).
- Prevalence: 10% general pop; 25% in students; and 42% in schizophrenics.

NatO: "Double Vision"

- Hypo metabolism in brain.
- Uncertain of how to describe symptoms
- Used the term, "double-vision."
- Non-standard double-vision: first, knows that a mental state exists (e.g. visual experience of a bird); *only after that*, does he become aware that the state belongs to self.

NatO: Somatoparaphrenia

- Feeling that one's limb belongs to someone else.
- Approximately 5% of RH stroke patients.
- The case of FB: somatoparaphrenia co-morbid with tactile extinction and neglect.
- Recovery of tactile sensation, by misrepresenting ownership.
- Only when touch "niece's" hand did FB feel the sensation.

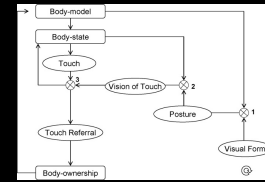
ExpP: Rubber Hand Illusion (RHI)

- Watching a rubber hand being stroked synchronously with one's own unseen hand causes the rubber hand to be attributed to one's own body, to "feel like it's my hand." (Botvinick & Cohen, 1998).



ExpP: RHI Empirical Model

- Tsakiris (2009) proposes a neurocognitive model (NCM) of body ownership that attempts to accommodate research on the rubber hand illusion (RHI).
- Phenomena: proprioceptive drift, touch referral, and hand ownership/disownership.



ExpP: RHI Extension to Full Body : Body-Swap Illusion (Petkova and Ehrsson, 2008)



ExpP: Body swap illusion

- Experimenter wears helmet with two closed-circuit television cameras (CCTV).
- The scenes that CCTV registered presented the experimenter's viewpoint.
- Subject wears head-mounted displays (HMD) and stands face to face with the experimenter.
- The subject's HMDs are connected to the experimenter's CCTV cameras such that the images from the CCTV were presented to the HMDs.



ExpP: Body swap illusion

- Effect: shift of perspective—the subject visually perceive himself rather than experimenter.
- Subject—adopting the experimenter's perspective—sees his own body, from shoulders to knees.
- Both extend hands, take hold, and squeeze.
- Control Condition: asynchronous squeezing.
- Illusion Condition: synchronous squeezing.



ExpP: Body Swap Illusion

- 20 subjects participated in this experiment, and each one was interviewed immediately afterwards.
- Participants' subjective experience: "after the experiment, several of the participants spontaneously remarked: ... 'I was shaking hands with myself!' (2008, 5).
- The subjects *misrepresent themselves* as squeezing their own hands.

ExpP: Out-of-Body Experience (OBE)

Video Ergo Sum: Manipulating Bodily Self-Consciousness

Basic Implications: "The Self" Hasno Boundaries?

Along with the growing use of virtual reality in research, the use of virtual reality in the study of bodily self-consciousness is also growing. In the last few years, the use of virtual reality in the study of bodily self-consciousness has been the focus of many studies. The use of virtual reality in the study of bodily self-consciousness has been the focus of many studies. The use of virtual reality in the study of bodily self-consciousness has been the focus of many studies.

ExpP: OBE

- Participants view back of own body filmed from 2m and projected onto HMB.
- Stoked for one minute, synchronously or (with time lag) asynchronously.
- Displace blindfolded participants and ask that they return to initial position.
- Modified RHI questionnaires—virtual body (or fake body) seems to be theirs.

Current Consensus

- The Tsakiris Model
- Anomalous cases—need not look like a real hand (or a real body).
- Body/Mental Ownership—no distinction.
- Is this one phenomenon or many?
- Getting clear about this is essential to explaining it.
- Proceed by pincers maneuver.

Mental Ownership Hypothesis (MOH)

- Two negative theses.
- Two positive corollaries.
- Seven conjectures.

MOH

- Introspective access to a mental state does not guarantee ownership.
- We deny: "it would be absurd to ask of someone who reports having a toothache 'are you sure it is *you* who have pains?'"
- Sometimes should ask W-S questions.
- Example: Somatoparaphrenia, Body Swap, Double-Vision, etc.

MOH

- There is nothing qualitatively special about introspection.
- We deny: "how we experience our body as *ours*... cannot be an external perceptual phenomenon. For when we perceive it from the outside, our body has no indelible stamp of ownership."
- Sometimes, e.g. RHI or Body Swap, perception of the external world determines ownership.

Appendix 2

MOH

- Conj One: self is minimalist—e.g. not autobiographical.
- Conj Two: mental ownership concerns relationship between self and psychological states, not self and body, per se. Dissociable, e.g. Moro et al. 2004.
- Conj Three: self an important theoretical construct. Example: harm to body differs from harm to self.

MOH

- Conj Four: Self is singular—a principal constraining factor in phenomena considered here is preservation of the sense that self is singular.
- Neural conflict—brain's need to integrate multi-modal stimuli—will, in most cases, be resolved so to preserve this singularity.

MOH

- Conj Five: belongingness/ownership is a distinctive form of representation.
- Sensory states can be represented—e.g. sound or pain when in deep sleep or vegetative states—without being re-represented as belonging to self. Recall the case of “double-vision.”
- Re-representation can involve spreading of activation from posterior cerebrum to frontoparietal areas.

MOH

- Conj Five Continued: various studies of “self-in-the-brain” suggest that ownership representations realized in virtue of multiple factors.
- Candidate factors: visual perspective, emotion, attention, vividness of imagery, empathy, etc.
- Example, case of empathy: feeling disgust in self and recognizing disgust in others—activity in anterior insula.

MOH

- Conj Six: RHI and Full-Body Illusions a good context for studying ownership-representations.
- Brain is forced to choose—where am I being touched, which is my hand, where is my body, is this my sensation, etc?
- Creation of a breach wherein ownership-representations reconsolidated.
- Preserve singular self.

MOH

- Conj Seven: ownership representations realized in virtue of patterns of functional and anatomical connectivity.
- These factors interact in ways that are poorly understood.
- But experimental paradigms of RHI and Full-Body Illusions afford opportunity to tease apart various components.

Our Experimental Probes

- First, test tolerance of boundary conditions.
- Consensus view seems to presuppose necessity of body-likeness.
- But mental-body ownership distinction does not.
- Check malleability, e.g. learning transfer and emotion.

Experimental Probes

- Second, check other correlations relevant to formation of ownership-representations.
- Empathy, dissociative personality, mental imagery, etc.
- Additional motivation: these other phenomena not well-understood. We think MOH can help promote integrated approach.

Experimental Probes

- Attention and Awareness: about one-third of subjects experience no illusion.
- Dual-task that calls for divided attention.
- Help to understand necessary conditions for development of ownership-representations.
- Help with teasing apart representation-of stimuli from representation-stimuli-as- belonging-to-self.

Experimental Probes

- Proper characterization essential.
- Possibly conflating distinct phenomena.
- SO, vary technique and stimuli (e.g. use temperature and pain)
- Employ psychometric techniques that include MOH-motivated W-S questions: e.g. for RHI, "are you sure it is *you* who felt the touch on the rubber hand?"

Experimental Probes

- Neural Substrates: MOH predicts that, to a degree, ownership-representations will be realized in the brain in diverse ways.
- Employ transcranial magnetic stimulation (TMS), diffusion tensor imaging (DTI), near infrared spectroscopy, etc. to specific areas (e.g. TMS applied to TPJ).
- Test "Shared Mechanism" vs. "Distinct Mechanism" Hypotheses.

Practical Applications

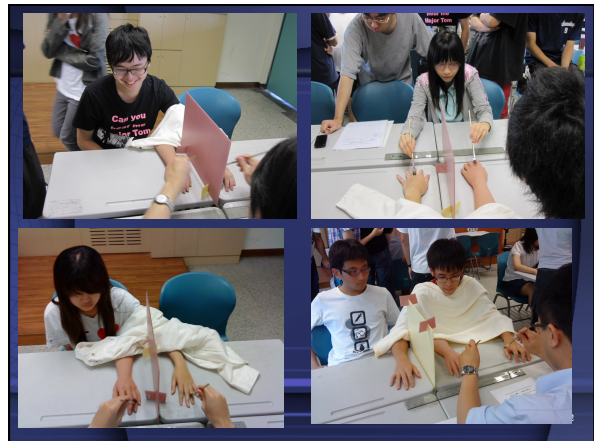
- By enhancing knowledge of malleability better able to predict and prepare for atypical conditions.
- For example: TPJ (as well as prefrontal) function altered by high altitudes and acute/chronic hypoxia; can cause OBEs.
- Enhanced use of prostheses as well as use of delicate tools in atypical environments.

Appendix 2

Conclusion

- Use MOH to motivate research concerning self, body, and ownership.
- Results motivate revision and refinement of MOH.
- Identify the critical psychological factors and neural mechanisms that determine ownership.
- MOH help to explain what now seem to be disparate phenomena.

Conference on Perception and Self-consciousness
Mental ownership and the rubber hand illusion
Timothy Lane
National Chengchi University
October 26, 2010



Our talk today

- 1. Two dubious assumptions and a working hypothesis
- 2. Pilot study in Taiwan
- 3. Philosophical and empirical implications
- 4. Neural mechanisms of Ownership
- 5. Conclusions

Part I. Dubious assumptions about self-consciousness

- First, body ownership is uniquely determined by introspection.
- Second, the Wittgenstein-Shoemaker (1968) question is absurd—"there is no question of recognizing a person when I say I have tooth-ache. To ask 'are you sure it is *you* who have pains?' would be nonsensical."

First Dubious Assumption

- E.G. Brewer (1995): "...how we experience our body *as ours*. Clearly this is not, and cannot be, an external perceptual phenomenon. For **when we perceive it from the outside, our body has no indelible stamp of ownership**. It appears just as one object among many..."

Problem with First Dubious Assumption

- Although "indelible stamp of ownership" is more likely associated with introspection than with external perception, **this is contingent**.
- What we introspect can feel alien, or fail to be represented as belonging to self: e.g. some cases of pain asymbolia.
- When we perceive others experiencing mental states, those states can (in a sense) be represented as belonging to self: e.g. vision-touch synesthesia
- More importantly: RHI and Full Body Illusions.

Second Dubious Assumption

- Our claim: The Wittgenstein-Shoemaker Question should be asked—‘are you sure it is you who is having this experience?’
- Why? Don’t subject and mental state fit together like “branch and branch bending?”
- **Somatoparaphrenia** (Bottini, 2002).
- **Body Swap**: “I was shaking hands with myself.” (Petkova & Ehrsson, 2008)
- **Atypical “double-vision”** (Zahn, et al, 2009)

7

More on “double-vision”

- Patient, healthy in all respects.
- But, visual perceptions distinct from ownership.
- Only detectable problem: hypo metabolism in right inferior temporal, parieto-occipital, etc.
- Knowledge of state, but no sense of belonging.

8

Take, as a working hypothesis

- The distinction between **Self-as-subject** and **Self-as-object** helps explain aspects of self-consciousness.
- Self-as-subject: I am the one who is in pain.
- Self-as-object: I am the one who is bleeding.

9

Part II. Phenomenology of RHI

- Conflicting tactile and visual stimuli lead to proprioceptive drift and touch referral.
- As well as hand ownership & disownership.
- Here, much of our focus on touch referral—“I feel tactile sensations on the rubber hand.”

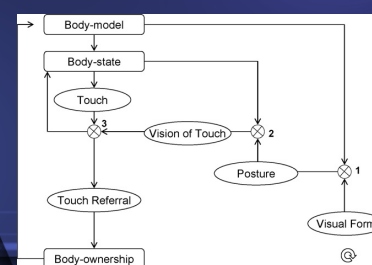
10

Phenomenology of RHI

- Psychometric approaches, e.g. PCA, to introspective reports (Longo et al. 2008).
- Attempt to evoke and quantify structures of experience.
- But what questions should be asked?
- In a sense, can only get out what you put in.

11

Tsakiris’ Empirical Model (2010)



12

Two concerns

- **Concern 1.** Longo et al. ask 27 questions, but neglect the distinction between **self-as-subject** and **self-as-object**. Their questions concern only **self-as-object**.
- **Concern 2.** **Body ownership** concerns **self-as-object**, whether a body part (e.g. a hand) or a full body belongs to me. **Mental ownership** concerns **self-as-subject**, whether I represent myself as the unique subject of experiences. **Body ownership** and **mental ownership** are distinct.

13

Two suggestions

- Longo et al. asked: "it seemed like the touch I felt was caused by the paintbrush touching the rubber hand." (2008, 983)
- Longo's question is about the cause, not about the sensation that is experienced.
- **Suggestion 1:** A subjective touch referral question should be asked: "I felt the touch on the rubber hand." Responses then judged on a Likert Scale: from +3 ("strongly agreed") to -3 ("strongly disagreed").

14

Our suggestions

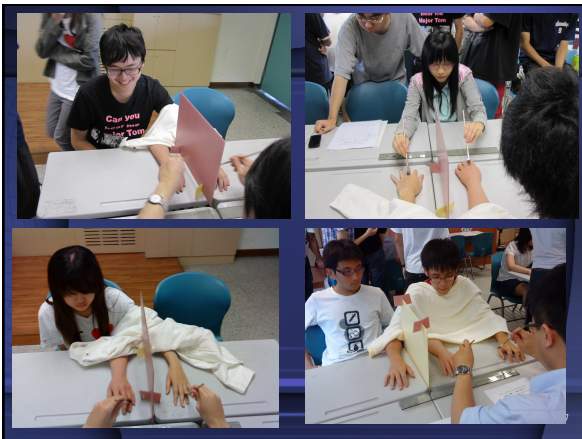
- **Suggestion 2:** Questions pertaining to **mental ownership** and its relation to body ownership should be asked.
- For example, "it seemed like **I was the one who** was feeling brush strokes on the rubber hand."
 - ◆ "I felt that the rubber hand was mine **before** I felt that the touch on the rubber hand was mine."
 - ◆ "I felt that the touch on the rubber hand was mine **before** I felt that the rubber hand was mine."

15

Our Predictions

1. Only after one approaches "strongly agree" (+2 or +3) on **subjective touch referral** will rubber hand ownership be experienced.
2. **Mental ownership**, ceteris paribus, is a prerequisite for **body ownership**.

16



Pilot study result 1 (only +2 or +3):

- 20 out of 66 subjects reported **body ownership** of rubber hand.
- 29 out of 66 subjects reported experiencing **touch referral**.
- 22 out of 66 subjects reported **mental ownership** of tactile sensations on the rubber hand.

18

Pilot study result 1:

- As predicted, **body ownership** only obtains when experiencing strong (+2 or +3) sensation of **touch referral**. (Only four anomalous cases.)
- As predicted, **mental ownership** correlates with **body ownership**. (Only one anomalous case.)

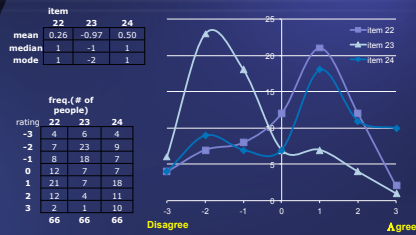
19

Pilot study result 2:

- As we predicted, **mental ownership precedes body ownership**.
- It seems to be the case that **mental ownership is a constraint on body ownership**.

20

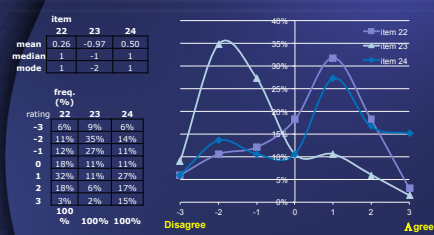
22. 在實驗過程之中，我覺得自己像個旁觀者在觀察自己的感受
 23. 我先感覺到那橡膠手臂是我的，才覺得那橡膠手臂被刷的觸覺是我的
 24. 我先感覺到橡膠手臂被刷的觸覺是我的，才覺得那橡膠手臂是我的



X axis: ratings, Y axis: # of people having that rating.

21

22. 在實驗過程之中，我覺得自己像個旁觀者在觀察自己的感受
 23. 我先感覺到那橡膠手臂是我的，才覺得那橡膠手臂被刷的觸覺是我的
 24. 我先感覺到橡膠手臂被刷的觸覺是我的，才覺得那橡膠手臂是我的



X axis: ratings, Y axis: % of people having that rating.

22

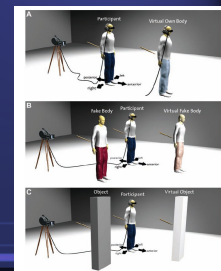
Part III. Philosophical and Empirical Implications

- The Self-Model Theory of Subjectivity (SMT) by Metzinger (2003, 2008, 2009).
- **“Mineness”**: “a higher order property of particular forms of phenomenal content ... an immediately given, non-conceptual sense of ownership.”
- Examples: “Subjectively, my leg is always experienced as being a part of me”; “My thoughts and feelings are always experienced as part of my own consciousness”; “My volitional acts are always initiated by myself.”

23

The Self-Model of Subjectivity (SMT) applied to the full-body illusion

- Metzinger regards self-identification as identification of a virtual body as one’s own body. **Self-identification** is taken to be “a globalized form of **identification with the body as a whole** (as opposed to ownership for body parts)” Metzinger and Blanke (2008, p. 8).



24

Our concerns

- Like Longo et al. (2008), Metzinger does not clearly distinguish self-as-subject from self-as-object.
- Our pilot study suggests that the distinction should be made.
- To understand self-consciousness there is a need to consider self-as-subject.

25

An empirical issue

- Distinct Mechanism (Metzinger and Blanke 2008, p. 9): “neural mechanisms of partial ownership and self-identification are...likely to differ.”
- Shared Mechanism (Tsakiris 2009, p. 9): “...the necessary conditions for the experience of ownership over a body-part seem to be the same as the ones involved in the experience of ownership over full bodies...available empirical findings from the two domains suggest that very similar neurocognitive processes are involved in ownership of body-parts and bodies.”

26

Our current position

- On this issue, we side with the Shared Mechanism Thesis (albeit with qualifications.)
- Why? From the perspective of mental ownership the distinction between body-part and full-body is not so important.
- What matters is that the sense of who is the subject of synchronous sensations remains constant.

27

Our current position

- Note that even for full-body illusions, only a small portion of the back is being stroked. What matters, what is critical to “self-identification” (where self is understood as subject, not object), is that the subject of synchronous stroking remains constant (or remains singular).
- Burden of proof resides with advocates of Distinct Mechanism Thesis.

28

Part IV. Mechanism of Self-as-Subject—Initial Thoughts

- Most discussions seem only concerned with self-as-object. Adequate explanation of these phenomena will require investigation of the self-as-subject mechanism.
- **Candidate(s) 1.** Northoff and Bermpohl on “Cortical Midline Structure” (CMS), Baars on “the observing self,” and Raichle on “the default system,” etc:

29

CMS, Observing Subject, Default

- For example: CMS allegedly, “mediates ownership of experience.”
- What is the neural substrate in virtue of which ownership is realized?
- Simplifying: predominantly right hemisphere, especially the right prefrontal and parietal cortex.

30

Motivation for such a claim?

- Non-conscious visual words activate word-processing regions of visual cortex.
- When conscious though, trigger additional, widespread, activity in fronto-parietal regions.
- Likewise for pain and distinction between conscious and automatic skills.
- Complementary studies for auditory stimulation of primary auditory cortex in vegetative states and deep sleep.

31

Link to the Default System?

- Why is the default system sometimes thought of as the self system?
- When subjects told to rest, Default System metabolism is higher than when performing various cognitive tasks; hypo-metabolism when unconscious.
- Perhaps spontaneous thoughts more self-relevant.

32

Problems?

- Fails to adequately distinguish self-relatedness from other-mind reading.
- "...the main brain regions recruited for others' mind representation are...the main brain regions reported in self studies..." (Legrand and Ruby 2009)
- L & R: an evaluation network (making inferences and recalling memories).
- Example: compare Repts of two faces.

33

L & R Alternative: "Self-Specific"

- "Self-Specific" –not "Self-Related"
- Experientially, we do distinguish self from non-self.
- A "perspective" grounds every perception and representation had by subject. E.G.: "My experience of tasting a lemon."

34

Self-Specificity

- L & R seek to characterize self-specificity in terms of perspective.
- Need to satisfy two (operational) conditions: "exclusivity" and "noncontingency."

35

Self-Specific

- Exclusivity: "a given self S is constituted by a self-specific component C only if C characterizes S exclusively."
- Noncontingency: "changing or losing C would amount to changing or losing the distinction between self and non-self."

36

So, does the self-specific idea help?

- For R&L the self-specific component is perspective.
- Recall their example: “My experience of tasting a lemon.”
- Exclusive and noncontingent?
- Perhaps.

37

Now recall “double-vision”

- Does self-specificity help capture the self-nonsel self distinction?
- To begin: Perspective does seem exclusive to the patient and a change of perspective would matter.
- But, at least for this patient, perspective doesn’t capture “mineness”—it’s not my visual state.

38

Where does this leave us?

- Recall that L & R concerned about failure of CMS to distinguish self from other.
- Consider “seeing” disgust on someone’s face and feeling disgust.
- Suppose anterior insula equally activated in both.

39

Where does this leave us?

- Humans do reliably, but fallibly, distinguish between the mental state’s of self and other.
- Simplifying, disgust representations in the anterior insula are, in turn, re-represented as belonging to self or other.

40

What is needed?

- Formal study of belongingness- or ownership representations.
- Provisionally (and crudely)—not wholly unlike CMS—sensory representations in posterior cerebrum, then forward spread of activation where parietal cortex re-represents (e.g. as allo- or ego-centric). Then, prefrontal.....

41

Tsakiris, Once Again

- Tsakiris on RHI: posterior parietal cortex (PPC) involved in resolution of conflict between visual and tactile, as well as resulting recalibration.
- But this is not enough: something more needs to be said about the mechanism in virtue of which ownership-representations (O-R) are formed.

42

Ownership-Representation (OR) Model

- ORs realized in virtue of patterns of functional and anatomical connectivity.
- Can be understood in terms of a hyper-region in an n-dimensional space, where “n” indicates the various neural substrates that enable instantiation of those mental states which contribute to the sense of belonging.

43

What does this imply?

- Ownership might well be more malleable than is now thought.
- E.G.: Failure to look like a hand might not be so important as is thought.
- Armel and Ramachandran results with distant objects or with assimilating tables into one’s body image.

44

Other factors involved in OR?

- Learning transfer.
- Emotion.
- Mental Imagery?
- Dissociative Personality?
- Etc.

45

Personal and Subpersonal Levels

- Subpersonal-Level Activity: vision more reliable and spatially acute than proprioception, so brain favors visual information.
- Personal-Level Constraint: singular (sense of) self is preserved.

46

RHI and Full-Body Redux

- If we are correct conflict resolution (between touch and vision) is resolved by degree, as ORs reconfigured.
- Creates an interim wherein it makes sense to ask W-S questions.
- A transition that allows for instability or uncertainty.

47

Conclusions

1. W-S Question should be asked.
2. Self-as-subject as working hypothesis.
3. Pilot study supports **mental ownership as a constraint on body ownership**.
4. Qualified Support for Distinct Mechanism Thesis.
5. CMS and Self-Specificity Insufficient.
6. Need to investigate the mechanisms that underlies self-as-subject or ownership-representations.

48

Media Impact on Individual Suicidality

A proposal for an ethical neuroimaging study

Kevin Chien-Chang Wu, MD, LL.M, PhD
National Taiwan University
Timothy Joseph Lane, PhD
National Chengchi University
Conference on Neuroscience, mind and the law Nov. 19, 2010

Presentation Outlines

- ◆ Suicide: definitions and history
- ◆ Suicide as a public health problem
- ◆ Justification of suicide prevention
 - ◆ Media reports and imitative suicide
 - ◆ Restricting freedom of speech and the principle of proportionality
- ◆ Randomized controlled trials: Psychological assessment and neuroscientific studies of media reports and individual suicidality
 - ◆ Analogy of SSRI experiments and ethicality analysis
- ◆ Preliminary conclusions

The fundamental question

There is but one truly serious philosophical problem and that is suicide. Judging **whether life is or is not worth living** amounts to answering the fundamental question of philosophy.

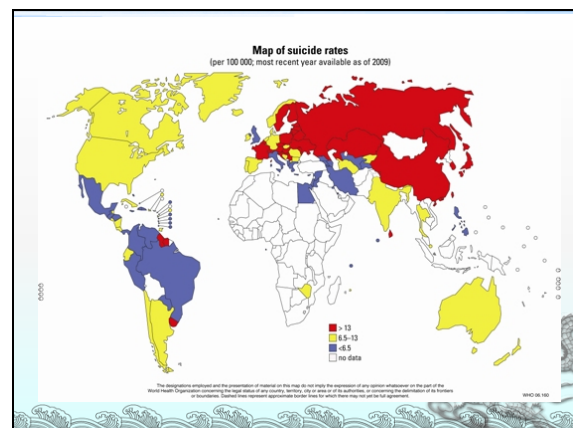
Camus

Definitions of suicide

- ◆ Durkheim: The term suicide is applied to all cases of death resulting directly from a **positive or negative act** of the victim, which he **knows** will produce this result
- ◆ Beauchamp's definition of suicide
 - ◆ A person intentionally bring about her own death
 - ◆ Not coerced
 - ◆ Death caused by conditions arranged by the person to bring about her own death

Suicide in history

- ◆ Ancient Greece and Rome
 - ◆ Suicide was allowed, and even praised by the stoics
- ◆ In middle-age Europe
 - ◆ Suicide was condemned and even punishable
- ◆ Around 18th century
 - ◆ Suicide as insanity
 - ◆ Decriminalization of suicide
- ◆ 19th century through now
 - ◆ Standard view: some suicides are not due to mental illness
 - ◆ Durkheim: suicide as social phenomenon – social determinants of suicide
 - ◆ Medicalization of suicide: diagnosis, treatment, prevention
 - ◆ Suicide as a public health problem

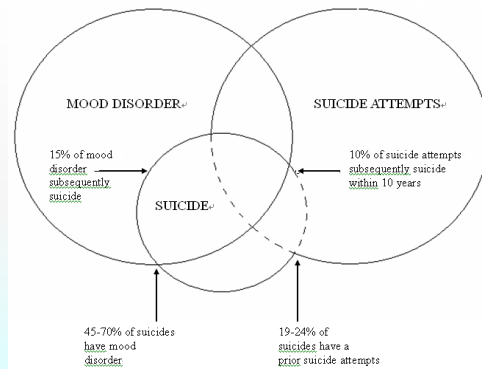
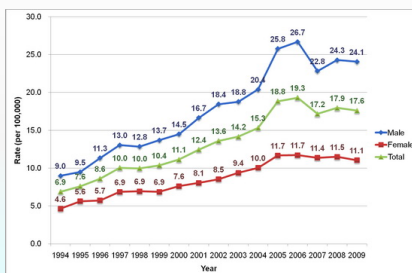


Appendix 4

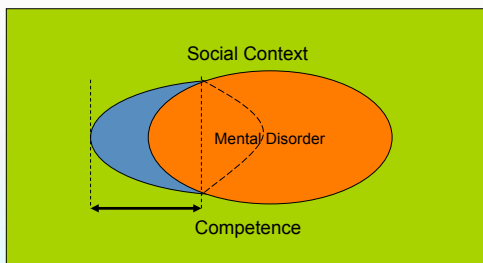
Suicide by the numbers
About 850,000 people commit suicide each year

Suicide has been ranked as the 9th leading cause of death in Taiwan for more than a decade

Trend of suicide in Taiwan



Suicide, mental disorder and competence

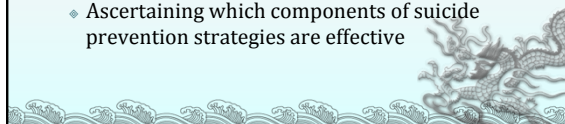


Determinants of suicide

- ◆ Biological factors
- ◆ Psychological factors
 - ◆ Psychoanalysis theories
 - ◆ Personality theories
 - ◆ Psychological stresses
 - ◆ Mental illness, etc.
- ◆ Social factors
 - ◆ Availability of lethal methods, unemployment, economic downturn, anomie, peer influence, **media report**

**Suicide prevention strategies:
A systematic review** (Mann et al, 2005)

- ◆ Two evidence-based effective strategies:
 - ◆ Physician education in depression recognition and treatment
 - ◆ Restricting access to lethal methods
- ◆ For optimizing the efficiency of resource utilization
 - ◆ Ascertaining which components of suicide prevention strategies are effective



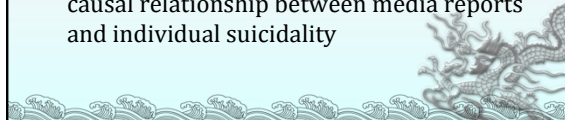
Justification of state power

- ◆ Social contract
 - ◆ Protection of individual freedom, rights, safety and welfare
 - ◆ Maintain public safety, interests and order



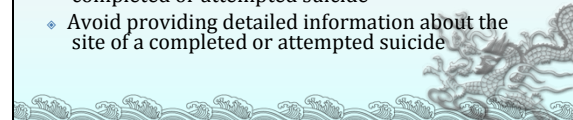
Science evidence of Media and imitative suicide

- ◆ Over 50 investigations into imitative suicides have shown that “media reporting of suicide can lead to imitative suicidal behavior” (WHO, 2008)
- ◆ However, most of them are ecological studies. So it is difficult to tease out the causal relationship between media reports and individual suicidality



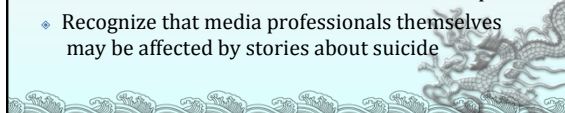
WHO quick reference guide for media professionals (1/2)

- ◆ Take the opportunity to educate the public about suicide
- ◆ Avoid language which sensationalizes or normalizes suicide, or presents it as a solution to problems
- ◆ Avoid prominent placement and undue repetition of stories about suicide
- ◆ Avoid explicit description of the method used in a completed or attempted suicide
- ◆ Avoid providing detailed information about the site of a completed or attempted suicide



WHO quick reference guide for media professionals (2/2)

- ◆ Word headlines carefully
- ◆ Exercise caution in using photographs or video footage
- ◆ Take particular care in reporting celebrity suicides
- ◆ Show due consideration for people bereaved by suicide
- ◆ Provide information about where to seek help
- ◆ Recognize that media professionals themselves may be affected by stories about suicide

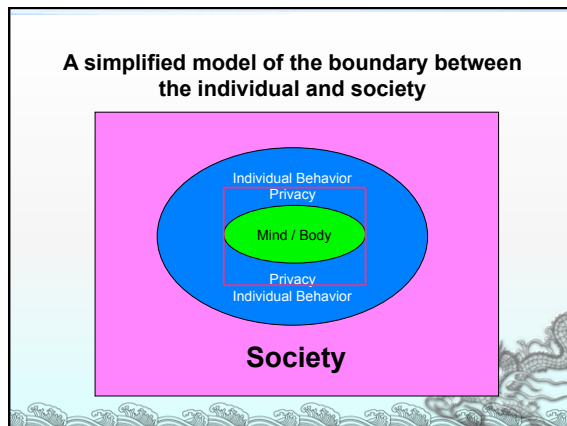


Do we want to regulate media reports of suicide by law?

Issues of freedom of speech



Appendix 4



Restricting freedom of speech:
Principle of proportionality

- ◆ Reasonable link between means and ends
- ◆ Least restrictive methods
- ◆ Benefits outweighing costs or risks

More scientific evidence

- ◆ Randomized controlled trials for assessing the impacts of media suicide reports on individual suicidality
 - ◆ Psychological levels by interviews and psychological tests
 - ◆ Brain function levels by neuroimaging: (1) seeing is believing; (2) finding biomarkers for suicidality
- ◆ Settings:
 - ◆ Subjects: mild to moderate depressed patients
 - ◆ Experiment group: reading sensational suicide reports
 - ◆ Control group: reading media reports following the WHO guidelines
 - ◆ All subjects must be admitted in acute psychiatric wards
 - ◆ Before and after psychological tests and functional neuroimaging (may include event-related potentials)
 - ◆ Discharged only when suicidality markedly improved and with close follow-up

Unethical experiments?

- ◆ Ecological studies are good enough?
 - ◆ Unavoidable ecological fallacy
 - ◆ It is a policy choice...
- ◆ Creating undue suicide risks to research subjects?
 - ◆ Current ecological studies have shown that imitative suicide usually is a short-term phenomenon
 - ◆ Improper suicide reports abound in the subjects' daily lives in Taiwan
 - ◆ We have good enough protection measures
- ◆ Subjects not competent?
 - ◆ Many patients with depression are still competent to consent to research

Analogy to SSRI experiments

- ◆ SSRI (Serotonin Selective Reuptake Inhibitors)
 - ◆ Scientific controversy about SSRI-induced suicide
 - ◆ Many governments finally put black-box warning about the relationship between SSRI and suicidality
 - ◆ Silence on the causal relationship between SSRI intake and individual suicidality
 - ◆ Scientific controversy of the therapeutic effectiveness of SSRIs
- ◆ Conundrums in research ethics
 - ◆ SSRIs are still prescribed by many physicians
 - ◆ Will the research ethics committee ban further experiments just because of the above information? Maybe not!!
- ◆ How about our media and individual suicidality research?
 - ◆ We do have the potential benefit for both the subject groups and the public

Preliminary conclusions

- ◆ We need to test our arguments by submitting our research proposal to the research ethics committee
- ◆ If the proposal is passed, we will have to worry about the practical difficulties
 - ◆ How many subjects could we have in the long run?
 - ◆ Where could we find adequate funding for the study?

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：100年6月17日

計畫編號	NSC 97-2410-H-004-154-MY3		
計畫名稱	自我與現象意識(3/3)		
出國人員姓名	藍亭	服務機構及職稱	國立政治大學心智、大腦與學習中心 教授
會議時間	100年6月9日 至 100年6月12日	會議地點	日本京都
會議名稱	(中文)意識科學國際研究學會第十五屆國際年會 (英文)The 15th Annual Meeting of the Association for the Scientific Studies of Consciousness		
發表論文題目	(中文)評論 Self-Specificity 的實徵典範 (英文) Self-Specificity and Mineness		

一、參加會議經過

My principle involvement concerned my own presentation, "Self-Specificity and Mineness," which is summarized below* and my role as chair for the session "Theories and Models of Consciousness." The session I chaired included the following talks:

1. Dolphin consciousness and higher-order thought theories; 2.

Attention and the structure of consciousness; 3. The sensorimotor approach and higher-order representationalism; 4. Consciousness, intentionality, and naturalism; 5. Inner clock model and conscious judgments of duration; and, 6. Self-oscillator model of bistable perception explains percept stabilization and reversal rate characteristics with interrupted ambiguous stimuli.

二、與會心得 To me the single most important thing were the constructive criticisms of three or four participants to my presentation. The second most important thing was the sixth talk indicated above (given in the session that I chaired), which demonstrated the applicability of an engineering model toward explaining a distinct form of conscious experience. The third most important were several keynote lectures delivered by young scholars, especially Fiona Macpherson on “Cognitive penetration of colour experience.”

三、考察參觀活動(無是項活動者略)

四、建議

A major theme of this particular meeting was the investigation of meta-cognitive abilities in apes. I have also found it regrettable

that, to my knowledge, no one in Taiwan studies the “Formosa Macaque”

(“臺灣獼猴”). I believe that this is an as-yet, untapped resource for rich study of cognitive activity.

五、攜回資料名稱及內容

1. 大會手冊

2. Matsuzawa, T. et al. Eds. 2011 Cognitive Development in Chimpanzees. Springer.

六、其他

*Legrand and Ruby (2009) argue that neural investigations of self mislead. “Self-relatedness” studies (e.g. Northoff et al. 2006), for example, fail to distinguish self from nonself. Neural substrates identified by that paradigm are not “specific” to self. Proclaiming a paradigm shift aimed at capturing that which is “constitutive” of self, they argue that inquiry should focus on “subjective perspective”—the relation between perceiving-subject and perceived-object. This new paradigm’s target is the experiential level, self-as-subject, and the minimal capacity to distinguish self from nonself. Perspective is seen as pivotal to understanding *specificity*, or “mineness” (see also Legrand 2007 and Christoff et al. In Press). But their operational definition of self-specificity—exclusivity and noncontingency—fails to account for mineness, as is shown by various empirical examples (e.g. Gott et al. 1984 and Zahn et al. 2008), wherein both conditions are satisfied but for perceiving-subject perceived-object is not “mine”. Previously I (Lane and Liang Forthcoming) have argued that mode-of-access is inadequate to account for mineness; subjective perspective also seems lacking. These failures, the lack of a conspicuous positive phenomenology of mineness, the fact that its loss can correlate with either attenuation or enhancement of sensory experience, and other factors suggest that search for a unique constituent is misguided. Specific, enabling processes might be found (e.g. Northoff et al. In Press). But mineness likely results from multiple parameters that interact dynamically, creating distinct regions within a multi-dimensional space. Among other things, this implies that mineness is multiply realizable and that it comes by degree.

國科會補助計畫衍生研發成果推廣資料表

日期:2011/12/28

國科會補助計畫	計畫名稱: 自我與現象意識
	計畫主持人: 藍亭
	計畫編號: 97-2410-H-004-154-MY3 學門領域: 心靈哲學
無研發成果推廣資料	

97 年度專題研究計畫研究成果彙整表

計畫主持人：藍亭		計畫編號：97-2410-H-004-154-MY3					
計畫名稱：自我與現象意識							
成果項目		量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）	
		實際已達成數（被接受或已發表）	預期總達成數（含實際已達成數）	本計畫實際貢獻百分比			
國內	論文著作	期刊論文	7	7	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	13	13	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	5	5	100%	人次	
		博士生	2	2	100%		
博士後研究員		0	0	100%			
專任助理		0	0	100%			
國外	論文著作	期刊論文	5	5	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	4	4	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
博士後研究員		0	0	100%			
專任助理		0	0	100%			

<p style="text-align: center;">其他成果</p> <p>(無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p>	<p>1. Arranged for visits by the following scholars:</p> <p style="padding-left: 20px;">A. Owen Flanagan, Chaired Professor of Philosophy and Neurobiology, Duke University.</p> <p style="padding-left: 20px;">B. Patricia Churchland, Chaired Professor of Philosophy, University of California at San Diego.</p> <p style="padding-left: 20px;">C. Neil Levy, Principal Investigator, Florey Neuroscience Institutes, University of Melbourne.</p> <p style="padding-left: 20px;">D. Frederique de Vignemont, Visiting Research Faculty, Department of Philosophy, New York University.</p> <p style="padding-left: 20px;">E. Walter Sinnott-Armstrong, Chaired Professor of Philosophy, Duke University</p> <p>2. 2010-2011 年：國立政治大學學術研究獎</p>
---	---

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

Here I indicate only the principal, already published, contribution: 在心理學、神經科學尚未發展的六〇年代，哲學家要研究心理或心靈，僅能透過內省或純思考方式。美國哲學家 Sydney Shoemaker 在 1968 年提出「IEM (immunity to error through misidentification)」學說，他認為當「我」做為主體，只要透過內省來感知痛覺、觸覺等感官經驗，那必然是「我」的感受，不可能辨識錯誤。例如當一個人說自己牙痛的時候，我們並不會質疑對方「那真的是你的痛覺嗎？」此外，Shoemaker 主張此關係不僅維持我們的體感經驗，同時也作用於行動意識與視覺感知。

然而，本研究透過實際病例與實驗結果，證明此關係並非必然。以研究病患症狀為例，某些病患會將自己的肢體視為「他人的」，而產生主體與意識經驗的分離。舉例而言，某病患將自己的左手視為外甥女，實驗設計反覆觸碰其左手，當病患被告知其左手要被觸碰時，她表示並無感覺，只有當實驗者告知她「外甥女的左手要被觸碰了」，病患才有觸覺反應。此病例說明雖然「我」是主體，卻必須將意識經驗表徵為「他人的」，才能透過內省去體驗並恢復感知。

本研究以此成功推翻了長年以來廣為接受的 Shoemaker 學說 (IEM)，認為其學說僅能視為「假說」，並非任何情況下都能成立。本研究旨在以心理學、神經科學等研究方法來檢驗哲學問題，並期望達成以下目標：(1) 透過告知臨床醫生哪些問題應被問及，協助醫生更了解病患情況；(2) 設計實驗使我們更了解「自我」、「意識經驗」與「身體」之間的關係。