Images and Constructs: Can the Neural Correlates of Self be revealed through Radiological Analysis?

Imágenes y constructos: ¿pueden los correlatos neuronales del self ser revelados a través de análisis radiológico?

Reflexive

Stanley B. Klein[a,*,]

[a] Department *of Psychological and Brain Sciences, University of California, Santa Barbara, CA, United States.*

Article history:

Received: 20-08-2013
Revised: 16-09-2013
Accepted: 17-09-2013

ARTICLE INFO

ABSTRACT

In this paper I argue that radiological attempts to elucidate the properties of self -- an endeavor currently popular in the social neurosciences -- are fraught with conceptual difficulties. I first discuss several philosophical criteria that increase the chances we are posing the "right" questions to nature. I then discuss whether these criteria are met when empirical efforts are directed at one of the central constructs in the social sciences – the human self. In particular, I consider whether recent attempts to map the neural correlates of self and its assumed properties using brain scanning technology satisfy the conceptual conditions minimally required to ask well-formed, theoretically satisfying questions of nature. I conclude that much theoretical work remains to be done.

Key Words:
Philosophy of science, self, theory, social neuroscience, neuroimaging

RESUMEN

En este artículo argumento que los intentos radiológicos para esclarecer las propiedades del Yo - esfuerzo actualmente popular en las neurociencias sociales – están plagados de dificultades conceptuales. Primero discuto distintos criterios filosóficos que incrementan las probabilidades de hacer las preguntas "correctas" a la naturaleza. Luego discuto si estos criterios son encontrados cuando los esfuerzos empíricos están dirigidos a uno de los constructos centrales de las neurociencias sociales – el Yo. En particular, tomo en cuenta si los intentos recientes de mapear las correlaciones neuronales del Yo y sus supuestas propiedades usando tecnología de escaneo cerebral satisfacen las condiciones conceptuales mínimamente requeridas para hacer preguntas bien formuladas y teóricamente satisfactorias de naturaleza. Concluyo que mucho trabajo teórico necesita ser realizado.

Palabras Clave:
Filosofía de la ciencia, Yo, teoría, neurociencia social, neuroimagen.

***Corresponding author:** S. Klein, Department of Psychological and Brain Sciences, 551 Ucen Road, University of California, Santa Barbara, CA 93106; Email: klein@psych.ucsb.edu.*

| ISSN printed 2011-2084 | ISSN electronic 2011-2079 |

1. INTRODUCTION

In 1993 John Kihlstrom, Mike Gazzaniga and I organized a symposium at the annual meeting of the Society for Experimental Social Psychology. Our goal was to introduce a sampling of techniques from the emerging field of neuroscience and show how their adoption might help social and personality psychologists obtain a better understanding of the issues they grappled with – e.g., self, person memory, attribution theory, etc. (e.g., Klein & Kihlstrom, 1998). Given the eclectic nature of our presentation, neither Kihlstrom nor I anticipated that shortly following its inception, social neuroscience would be co-opted by one particular methodology – radiological measures of neural structure and function. In the early days of social neuroscience, the image held sway. What was not so clear was whether this was a good thing – a point I elaborate in this article.

Fortunately, after a decade's infatuation, this methodological myopia has begun to remit. A look through the table of contents of any recent volume of *Social Cognitive Neuroscience or Social, Cognitive and Affective Neuroscience* (the two primary outlets for social neuroscience) reveals a diversity of methodologies including, but not limited to, neuroimaging techniques, electrophysiological recording, brain lesion studies, and hormonal assays. This is a good thing – it is what a neuroscience of social and personality psychology should be (e.g., Klein, Lax, & Gangi, 2010). But fMRI (the technological successor to PET, CAT and SPECT scans) still occupies a position of methodological dominance (as it also does in the cognitive neurosciences).

In this article I call to question what data obtained from radiological measures allows us to conclude about theoretical constructs of interest in social and personality psychology. Specifically, building on concerns voiced by Uttal (2001), I argue that many of the constructs in the social neurosciences are seriously under-specified with regard to both the observational implications and conceptual aspects of theory. Accordingly, radiological attempts to map the relation between those constructs and their "neural correlates" suffer from conceptual ambiguity in addition to interpretive concerns that have been raised about the methodology, per se (e.g., Coltheart, 2006; Dumit, 2004; Klein, 2010-a; Savoy, 2005; Uttal, 2001). Put differently, in the domain of neuroimaging, attempts to map observation to theory (so-called "correspondence rules"; e.g., Carnap, 1936) are plagued by serious issues that attenuate the force of the conclusions we can draw about the relation between psychological constructs and their neural instantiation (note: these concerns also apply to the use of radiological tools in the cognitive neurosciences).

In this paper I focus on the "theory" side of the relational mapping. Readers interested in issues pertaining to methodology are referred to Dumit (2004) for a non-technical overview. To cut to the chase, conceptual under-specification characterizes many of the theory-based targets that radiological evidence is marshaled to address. This makes attempts to map the relation between construct and image an enterprise fraught with ambiguity (e.g., Klein & Gangi, 2010; Uttal, 2001). The scientific merit of neuroimaging analyses of constructs rests on the assumption –first voiced by Plato in his famous dialogue, Meno (e.g., Day, 1994)– that the construct (receiving radiological attention) cuts nature at its joints. However, just because a scientist has a particular way of dividing the conceptual landscape does not guarantee that nature recognizes the theoretical distinctions he or she proposes.

While there are many constructs in social and personality psychology I could discuss with regard to the issue just raised, I focus on one with which I have the most empirical and conceptual familiarity – the self. No less an authority than William James argued that the self is the *fundamental* unit of analysis for a science of mental life, the problem about which everything else revolves (1890). It would be hard to find a more psychologically propitious construct with which to illustrate my points.

2. TESTING SCIENTIFIC THEORIES – THE ART OF ASKING THE "RIGHT" QUESTION

The scientific approach to understanding consists in asking questions of nature in the service of examining some provision or predicted consequence of a conceptually well-specified theory (e.g., Godfrey-Smith, 2003; Hanson, 1958; Klee, 1997; Ladyman, 2002; Trusted, 1979). This questioning most often is accomplished by means of

experimentation.[1] It is essential to keep in mind, however, that the scientific enterprise is *more* than the construction and interpretation of the observational offerings of well-designed experiments. As Heisenberg (1999) notes: "What we observe is not nature itself but nature exposed to our method of questioning" (p. 58) accordingly, asking the "right" question is critical to obtaining a solution to problems posed to nature. Add to these considerations the truism that "one has a better chance of finding what he or she seeks if one knows what it is he or she is looking for" and we have a concise statement of the basic considerations of scientific inquiry -- as well as its potential pitfalls. Questions need to be posed. But not just any questions – the "right" questions must be asked.

And therein lays the rub. The idea that if we ask enough questions, nature will reveal her "truths" -- an approach sometimes called the naïve Baconian empiricism – misrepresents of how science actually, or is supposed to, work (e.g., Klee, 1997; Newell, 1973). Not any question will suffice: We need to pose the "right" questions to nature to distinguish fact from conjecture; that is, to cut nature at its seams.

But, how do we know what constitutes the "right" question? The obvious answer is that theory guides our selection: A well-formulated theory grounds our empiricism by specifying what *it is* that we are looking for, as well as providing the experimental conditions necessary to ensure that the answers nature provides correspond to the questions posed (e.g., Brunswik, 1947/1956; Godfrey-Smith, 2003; Trusted, 1979).[2] Unfortunately, theoretical formalizations of the "*its* of interest" in the social neurosciences often are too under-specified to enable us to ascertain whether the questions we have asked nature are the "right" ones (e.g., Klein, 2013; Newell, 1973).

The field of psychology is awash in data. What too often is lacking are overarching, nomologically sophisticated theories by which that data can attain conceptual relevance. Consequently, the *it* under analysis more often is stipulated than theoretically specified (e.g., Klein, 2013). Wittgenstein frames the problem in the following manner:

> The confusion and barrenness of psychology is not to be explained by its being a 'young science'; its state is not comparable with that of physics, for instance, in its beginnings…for in psychology, there are experimental methods and conceptual confusion. The existence of the experimental method makes us think that we have the means of getting rid of the problems which trouble us; but problem and method pass one another by. (Wittgenstein, 1953/2009, A Fragment XIV, 371; emphasis in original).

Fodor argues along similar lines:

> Psychological metatheory has remained seriously underdeveloped…a psychologist is likely to appeal his decisions about research strategies directly to general methodological principles to an extent to which a physicist of chemist does not…a consequence of the unsettled state of psychological metatheory is thus that schools of psychology are distinguished as much by the kinds of experiments that their adherents typically perform as by the theories they espouse. (Fodor, 1968, p. xiv-xv).[3]

In short, the social sciences suffer from a paucity of abstract formalizations capable of supporting the type of precise theory-based predictions and specifications found in most of the physical sciences (e.g., Klee, 1997; Klein, 2013; Ladyman, 2002; Trusted, 1979; Wittgenstein, 1953/2009). Absent a sophisticated network of nomological interdependencies between observation and theory (e.g., Margenau, 1950; Torgerson, 1958), social scientists are at a clear disadvantage when it

[1] The role of the experimental method in science often is over-simplified. Contrary to popular belief (both lay and professional), there are no definitive experimental tests of the predictions of a theory, nor can any theories, by themselves, permit the formulation of specific observational predictions. Such predictions require the theory under consideration be conjoined with a host of additional, interdependent assumptions, beliefs, guesses, metaphysical commitments as well as other theories (e.g., Hanson, 1958; Klee, 1997).

[2] Questions sometimes are asked in advance of the establishment of theory -- for example, when questioning is more exploratory than in the service of Popperian falsification (i.e., the hypothetical-deductive method of scientific inquiry; e.g., Ladyman, 2002; Popper, 1963/2004). However, my concern is with empiricism in the service of testing existing theory rather than establishing it.

[3] These concerns were expressed, in part, as a response to the behaviorist influence on psychology. Things have changed (in my opinion, for the better) since the behaviorist approach ceased to be the dominant orientation of psychological investigation. However, the critiques offered by Wittgenstein and Fodor unfortunately still are relevant to many psychological research programs (e.g., Klee, 1997).

comes to determining whether our experimental outcomes are commensurate with, or antithetical to, whatever theoretical topic is under examination (e.g., Newell, 1973).

Take the self as our example. Its assumed properties and causal potencies typically are posited *absent* the conceptual treatment needed to ensure that our attempts at empirical explication ask the "right" questions (e.g., Klein, 2012a). As I will argue in the next section, conceptual confusion about what *it* is that we refer to by the word "self" makes it a priori impossible to know whether the answers nature offers in response to radiological interrogation provide insight into the construct under scrutiny. As Wittgenstein implies in the quote cited above, our experimental procedures may be telling us more about the task performed than the construct it is recruited to investigate.

3. THE SELF AS THE OBJECT OF SCIENTIFIC INQUIRY

The self is perhaps the most familiar yet elusive aspect of human experience. It has captured the imagination of investigators and theoreticians from a vast and diverse array of academic disciplines and cultural traditions. For example, psychological treatments can be found in Allport (1961), Conway (2005), Freud (1949), Gergen (1971), James (1890), S. Klein (2001, 2010, 2012a), Leahy (1985), Leary and Tangney (2012), Lecky (1945), Legrand and Ruby (2009), Neisser (1988), Symonds (1951); philosophical treatments (both Western and Eastern) include Albahari (2006), Chadha (2013), Dainton (2008), Ganeri (2012), Ismael (2007), Jopling (2000), Lund (2005), Siderits (2003), Strawson (2009), Williams (1973), Yao (2005), Zahavi (2005); literary analyses are provided by Bruner (2002), Eakin (2008), Freeman (1991), King (2000), Parker (2007), and so on. As the reader likely is aware, this listing barely scratches the surface of the discourse and cross-disciplinary interest the self has attracted for more than 2,500 years (e.g., Sorabji, 2006). But the point has been made – the self is one of the most heavily investigated and hotly debated concepts in the history of recorded human thought.

So, what is a self? To what does the term refer? As both psychologists and philosophers have made abundantly clear, a theoretically compelling answer is elusive at best (for review, see Klein, 2012a). Klein and Gangi (2010) have proposed that Bertrand Russell's (1912/1999) distinction between knowledge by acquaintance and knowledge by description (which actually traces from the work of Augustine of Hippo in his Confessions) provides a perspective on the difficulties we encounter when attempting to translate our everyday experience of self into descriptively satisfying theoretical terms. Russell proposed that we have knowledge by acquaintance when we know something via direct personal contact (sensory or introspective) and exhibit that knowledge by using appropriately referential terms when we communicate with others. As regards the self, this is seen in the ease with which we talk about the self as well as understand talk about self by others.

But knowledge by acquaintance does not guarantee that we can capture that knowledge in descriptive form. When we attempt to make explicit what it is to which we refer – i.e., when we are asked to describe what the word "self" means — problems quickly arise. Despite more than 2,500 years devoted to the "problem of the self", it has proven notoriously difficult to provide a set of propositions capable of transforming our acquired knowledge into a description of what a self *is* (e.g., Klein, 2010, 2012a).

Some have argued that the reason there is a problem is that the question is based on a false premise -- the illusion that there is an elusive self to be explained (e.g., Dennett, 1991; Hood, 2012; Hume, 1739-1740/1978; Metzinger, 2009; Pessoa, Thompson, & Noe, 1998; for a critical discussion, see Siderits, Thompson, & Zahavi, 2011). On this view, the issue of addressing the "right" question to nature becomes moot. There is no question in need of an answer.

A difficulty with this idea that the self is an illusion is that an illusion is an experience, and an experience requires an experiencer (e.g., Klein, 2012a; 2012b, in press; Schwerin, 2012; Strawson, 2011; Zahavi, 2005). Accordingly, unless one wishes to maintain that an "experiencer" is something other than a self (which would be more a definitional stipulation than a conceptual entailment), this approach amounts to an exercise in begging the question. Meixner (2008) puts it bluntly, "The fictionalization of subjects of experience is incoherent, since it involves the incoherent idea that

I, for example, am an illusion of myself" (p. 162). Kant (1781/1998) goes further, arguing that the self of subjective awareness (his transcendental ego) *must* accompany experience (for related views, see James, 1890; Lund, 2005).

Despite concerns about the self's ontological status, psychologists have not shown any reluctance to put the term to work in an abundance of self-hyphenated compounds (e.g., self-comparison, self-concept, self-complexity, self-deception, self-esteem, self-handicapping, self-image, self-perception, self-regulation, self-reference, self-verification; for reviews see Kihlstrom et al., 1988; Leary & Tangney, 2012). But, the question remains: What is *it* that is being verified, conceptualized, complicated, esteemed, deceived, verified, regulated, and handicapped? Unfortunately, the focus of these research agendas rests firmly on the entries on the right-hand side of the hyphenated relation -- to the detriment of our understanding of the self in terms of its properties and causal potencies (for discussions see Klein, 2010, 2012a; Klein & Gangi, 2010).

Despite these ontological concerns (and other, more serious, ones discussed below), models of the self have been on display in psychology for over 100 years (e.g., Calkins, 1915; Conway, 2005; Greenwald, 1981; James, 1890; Kihlstrom & Klein, 1994; Neisser, 1988; Samsonovich & Nadel, 2005; Stuss, 1991). Yet, the elusive nature of the construct has resulted in most of these offerings concentrating on the task of explicating the "self" in its assumed causal or foundational relation to a specific set of predicates, processes and contexts (c.f., Leary & Tangney, 2012; Sedikides & Brewer, 2001). We thus find models of contextualized selves, cultural selves, social selves, cognitive selves, embodied selves, situational selves, autobiographical selves, relational selves, narrative selves, collective selves, etc. But consideration of what the self *is* that serves as the assumed bedrock for these cultural, social, cognitive and narrative instantiations has been vastly under-specified (for discussion, see S. Klein & Gangi, 2010; S. Klein, 2010, in press).

Conceptual difficulties surrounding the term "self" are not restricted to psychology. The "problem of the self" has been intensely debated in philosophy and theology (both Western and Eastern) for more than 25 centuries (for review, see Sorabji, 2006). The ongoing, multi-disciplinary nature of these (often contentious) examinations of the self (for discussions, see Baillie, 1993; Flanagan, 2002; Gallagher & Shear, 1999; Giles, 1997; Johnstone, 1970; Siderits et al., 2011; Vierkant, 2003) has left some wondering whether a conceptual understanding of self is possible in practice (e.g., Olson, 1999) or in principle (e.g., McGinn, 1991).

4. WHY THE DIFFICULTY?

One reason for many of the difficulties we face when attempting to describe what we mean by the word "self" is that there is *not a single* self to describe (e.g., Klein, 2001, 2004, 2010, 2012a, in press; Klein, Cosmides, Tooby, & Chance, 2002; Klein, Rozendal, & Cosmides, 2002; Legrand & Ruby, 2009; Neisser, 1988; Stern, 1985). Rather, *two* ideas of the self are involved in almost every discussion of the topic, although these ideas rarely are separated. On this analysis -- reviewed at length in S. Klein (2012a, in press) -- the self meaningfully can be partitioned into two distinct, but normally interacting, aspects – the neurally instantiated systems of self-knowledge and the self of first-person subjectivity (e.g., James, 1890; Legrand and Ruby, 2009; Strawson, 2009; Zahavi, 1999, 2005).

While it is well beyond the scope of this paper to go into detail about these two aspects of self (extensive discussion can be found in S. Klein, 2012a, in press), the reader can, if it helps, treat them as "roughly" analogous to James's (1890) self-as-known and self-as-knower (see Klein, in press, for the reason scare quotes bracket the word "roughly"). These two aspects of self cannot be deduced from, or reduced to, a single, underlying principle, structure, process, substance or system (e.g., Kant, 1781/1998; Klein, 2012a, in press; Zahavi, 2005).[4] One – the neuro-cognitive systems of the psycho-physical person (consisting of such things as personal memory, body image, emotions) – is materially (primarily neurally) instantiated and therefore capable of being apprehended and treated as an *object* of scientific inquiry. The other – the self of first-person

[4] Theoretical considerations and empirical findings that, I believe, offer strong support for the position that these two aspects or types of self are not reducible, one to the other -- that is, they are *not* different ways of thinking about a single entity – are presented in S. Klein (2012, in press).

subjectivity – is the *subject* having the experience, rather than the object of that experience. Logically, this entails that it cannot be *directly* known by acts of perception or introspection (e.g., Earle, 1972; Kant, 1781/1998; Klein, 2012a; Lund, 2005; Swinburne, 1997, 2013; for a different view, see Strawson, 2009). Our knowledge of the self of subjectivity is a matter of acquaintance or feeling, not something that can be treated as the *object* of descriptive analysis (e.g., Kant, 1781/1998; Klein, 2012a, in press; Nagel, 1974).

Despite differences in ontological status, under normal circumstances these two aspects of self interact, and this interaction is a prerequisite for our acquaintance with our self. Indeed, it is *only* via their interaction that a particular form of consciousness – self-awareness —becomes possible (these assertions are treated extensively in Klein, 2012a, in press). Following Fitche's dictum, there can be no subject without an object or an object without a subject (e.g., Neuhouser, 1990).

These arguments have important consequences for the "self" as an object of scientific inquiry. Considerable progress has been made describing the cognitive and neurological bases of the self of neuro-cognitive instantiation (e.g., Conway, 2005; Kihlstrom & Klein, 1994, 1997; Klein, 2010; Klein, Cosmides, Tooby, & Chance, 2002; Neisser, 1988; Samsonovich & Nadel, 2005). This is because the neuro-cognitive bases of self-knowledge can be (and have been) objectified, and thus are amenable to scientific analysis.

The self of first-person subjectivity, by contrast, is too poorly understood to bear the definitional weight required to stand in a meaningful causal relation between the self and its predicates (e.g., regulation, image, conception, complexity, handicapping, verification, etc.). Moreover, as discussed below, treating the subjective self as an object has the scientifically unhappy consequence of stripping it of its core feature – i.e., its subjectivity.

Not surprisingly, many researchers (intentionally or otherwise) avoid these difficulties by relying on readers' familiarity with the term "self" -- derived from years of knowledge by acquaintance -- to confer a (false) sense of confidence that he or she knows what *it is* to which the author refers. But the problem remains – in terms of a descriptive, theoretical analysis, we remain unclear what it is we refer to when we apply the label "self".

Compounding the problem, researchers often fail to appreciate that the self of subjectivity is *not* the object of their experimental studies (e.g., Klein, in press; Klein & Nelson, 2014). Nor could it be. Objectivity is based on the assumption that an act or object exists independent of *any* individual's awareness of it (e.g., Earle, 1955; Martin, 2008; Nagel, 1974; Rescher, 1997); it is something *other* than the self. When objectivity is the stance adopted by the self to study itself, the self must, of necessity, be directed toward what is not the self -- to some "other" that serves as the self's object (e.g., Earle, 1972; Husserl, 1964; Lund, 2005; Klein, 2012a, in press; Zahavi, 2005). Thus, to study myself as an object, I must transform myself into an "other", into a "not-self". Accordingly, the subjective self is not, and cannot, be an object for itself and still maintain its subjectivity. Paradoxically, the subjective aspect of self can achieve objectivity only at the cost of forfeiting its essence as a subjective center (e.g., Klein, 2012a). This remains the case whether investigation is conducted by first-person introspection or third person observation: Once apprehended by first-person subjectivity (one's own or that of another person), the subjective self becomes an object in the manner all objects (both mental and physical) must, of necessity, become when apprehended (e.g., Husserl, 1964). In the process, the subjective core of the self of first-person experience is lost.

These ideas have considerable resonance with the philosophical position, first voiced by 19th century philosopher Franz Brentano (1995) that consciousness is, of logical necessity, intentional. The term "intentionality", with regard to conscious experience, means that the consciousness must be *about* some object or idea *other* than itself -- it must have an object (whether physical or mental). Consciousness, absent its object, is opaque to experience whether that experience is felt, introspected or perceptually given (see also James, 1890; Kant, 1781/1998; Klein, 2012a; Lund, 2005; Zahavi, 2005). Returning to the self, its subjective aspect, devoid of an intentional object, cannot be apprehended regardless of whether that apprehension depends on personal acquaintance or logical inference.[5]

[5] It merits mention that various Eastern wisdom traditions take issue with this position, arguing it is possible, with sufficient training, to attain a state of "pure" consciousness (i.e., consciousness absent an object; for review, see Forman, 1990).

The subjectivity of self thus seems a poor candidate for scientific exploration (radiological or otherwise), an enterprise predicated on understanding the behavior and dispositions of objects in relation to underlying theory (e.g., Hanson, 1958; Klee, 1997; Ladyman, 2002; Quine, 1951). Science trades in the world of publically observable and physically measureable objects and events (e.g., Earle, 1955; Margenau, 1950; Rescher, 1997). Correspondence rules play an important role in this enterprise, providing the mechanism by which investigators can construct logically defensible mappings of observables to theory (e.g., Carnap, 1936). Despite attracting the critical attention of philosophers (e.g., Hempel, 1965), this approach continues to have considerable influence in the social sciences – psychology in particular (e.g., Klee, 1997). For the subjective self to become part of the scientific world it has to relinquish its subjectivity. But then we no longer are investigating the subjective properties our research has been designed to illuminate. To maintain its identity, the subjective self cannot be transformed into an object (e.g., Kant, 1781/1998; Klein, 2012a; Zahavi, 2005). And, absent an objective instantiation, the rules of correspondence cannot be applied in any straightforward manner, thus making it hard to see how scientifically defensible questions about this aspect of the self can be formulated.

In contrast to the subjective aspect of self, the neurally instantiated aspects can, in virtue of their materiality, be submitted to a scientifically viable descriptive analysis. However, even here problems exist, a number of which are detailed in S. Klein (in press). For our purposes, the key concern is that investigators too often fail to appreciate that the objective aspect of self admits to a multiplicity of constituents.

As one example, within the domain of self-referential memory, research has revealed at least three different types of self-knowledge that cut both across and within systems of long-term memory – i.e., episodic personal memories, semantic factual self-knowledge and semantic trait self-knowledge (for recent reviews, see Klein & Gangi, 2010; Klein & Lax, 2010; Martinelli, Sperduti, & Piolino, 2013; Picard et al., 2013; Prebble, Addis, & Tippett, in press; Renoult, Davidson, Palombo, Moscovitch, & Levine, 2012). These memory-based aspects of self are both conceptually and functionally independent (e.g., Klein & Gangi, 2010; Klein & Lax, 2010; Renoult et al., 2012). Consequently, one must take considerable care when attempting to map the memorial properties of the self onto its assumed neural correlates: It matters greatly *which* memory-based aspect of the self is hypothesized to mediate the phenomena (physical and/or mental) under investigation.

Moreover, it is well-known (e.g., Klein & Loftus, 1993; for review see Klein, Robertson, Gangi, & Loftus, 2008) that the same experimental outcome can be mediated by more than one system of self-memory. For example, judgments about a trait's self-relevance can be accomplished by either semantic *or* episodic memory, the contributions of each determined in large part by the amount of experience one has with the trait being judged (e.g., Klein, Loftus, Trafton, & Fuhrman, 1992). Further complicating matters, under certain conditions judgments about a trait's self–relevance can *simultaneously* activate episodic *and* semantic systems of self-knolwedge (Klein, Cosmides, Tooby, & Chance 2001, 2002). Unfortunately, most recent attempts at the radiological localization of trait knowledge (both self and other) are uniformed by such considerations (e.g., Ma et al., 2013).

Thus, conclusions based on radiological analyses of trait self-judgments can, and, as we will see below, often do provide multiple, non-overlapping localizations unless correlated factors (e.g., trait-relevant behavioral experience) are carefully controlled. Unfortunately, to the best of my knowledge, brain scanning studies of trait self-judgments – of which a number exist (e.g., Craik et al., 1999; Kelley et al., 2002; Macrae, Morgan, Heatheron, Banfiled & Kelley, 2004) -- have yet to take such precautions.

5. SUMMING UP THUS FAR

Asking the "right" of nature question requires we be clear about the properties (both observable and theoretical) of the concept toward which our questions are directed. This, in turn, depends on the availability of a well-formulated conceptual analysis of the *it* under investigation.

In the case of the "self", theoretical formalization often is lacking or seriously under-specified, leading researchers to rely, to an uncomfortable degree, on pre-theoretical

commitments (based largely on personal acquaintance) as a way of compensating for the lack of consensually shared, conceptually sophisticated theory (e.g., Klein, 2012a, 2013). Formulating the "right" questions to address to nature is particularly problematic when the *it* under investigation is not only theoretically under-specified (e.g., Klein, 2012a; Olson, 1999) but also considered (at least by some) to be existentially questionable (e.g., Hood, 2012; Hume, 1739-1740/1978; Metzinger, 2009).

Adding to these difficulties, one aspect of the self (its subjectivity) may not be amenable to description by a conceptually satisfying set of operational or theoretic propositions even if existential concerns are shown to be misguided. The problem is twofold. First, transforming the subject into an object of experience (a process required for its apprehension) has the effect of stripping the subject of its core subjectivity. The second, related, issue involves the intentionality of consciousness (e.g., Brentano, 1995; Textor, 2013). Applied to imaging studies, the intentionality of consciousness makes it logically impossible to ascertain whether the neural systems identified via radiological analysis of self-consciousness reflect the activity of (a) the self as subject, (b) the self as the object of subjectivity, (c) an object of subjectivity other than self, (d) all of the above, or (e) none of the above (these concerns apply to any attempt to identify the neural correlates of consciousness). As it will be detailed in the next section, radiological attempts to localize the neural substrates of self-reflection are especially problematic due to the inherent confound between the act of reflection and its object.

With regard to the neuro-cognitive aspects of self, a primary obstacle to radiological analysis concerns the failure of most investigators to provide reason to believe that they are familiar with the complexities of the systems under examination. As noted previously, at least six different, functionally independent systems of self-knowledge have been identified; and within those systems a variety of subsystems have been found – many of which also are functionally independent (for recent reviews, see Klein & Lax, 2010; Martinelli et al., 2013; Renoult et al., 2012).

At least this problem has a clear path to resolution. But until steps toward increased familiarity with the complexities of the neuro-cognitive self are evidenced, the conclusions of radiological analyses of the "self" continue to read more like mappings of "technology to terminology" than mappings of "technology to theory".

As if that were not enough cause for concern, it also must be remembered that some investigators contend that the self is an illusion (e.g., Hood, 2012; Metzinger, 2009), or, if it does exists, it is beyond our capacity to (descriptively) know (e.g., McGinn, 1991). Such sentiments do not bode well for radiological analyses -- it is logically impossible to neurally localize constructs that either lie outside our conceptual reach, or do not exist in the first place.

In sum, even if the *it* of our investigations is not an illusion, it certainly is elusive (e.g., Klein, in press; Lewis, 1982). In the following section I discuss some implications of the conceptual issues surrounding the concept of "self" for social neuroscience's attempts to map its neural substrates.

6. THE SELF AS REVEALED BY NEUROIMAGING

The first examination of self using brain scan technology was conducted by Craik et al. (1999). Since that beginning, the number of radiological investigations of the self has been so prolific that not long after Craik et al's (1999) initial foray it already made sense to conduct meta-analytic reviews of the accumulated data (e.g., Northoff et al., 2006). And, judging by the number of publications that have appeared in just the past 5 years, interest in radiological approaches to self shows no evidence of waning. Indeed, the literature has grown so explosively that there now exist specialized meta-analyses devoted to questions within sub-domains of self research (e.g., the meta-analysis of fMRI studies of self-other judgments; Denny, Kober, Wager, & Ocshner, 2012)

Rather than attempt an exhaustive review of this voluminous literature (good luck to anyone brave enough to take on that assignment!), I focus my comments on a sampling of papers purporting either to "locate the self in the brain" or to "examine the neurological basis of self-reflection". Although I just as easily could have examined radiological studies of the neural underpinnings of self-knowledge (e.g., Moran, Kelley, & Heatherton, 2013; Ochsner et al., 2005), self-referential thought (e.g., Abraham,

2013; Gusnard, Akbudak, Shulman, & Raichle, 2001; Leshikar & Duarte, 2013), the self-concept (e.g., Heatherton, Macrae, & Kelley, 2004; Kim & Johnson, 2010), self/other differences (e.g., Chen, Wagner, Kelley, Powers, & Heatherton, 2013), trait self-knowledge (e.g., Ma et al., 2013), self-regulation (e.g., Beauregard, Lévesque, & Bourgouin, 2001: Heatherton, 2011), etc., the two topics I address have the advantage of being *particularly* clear examples of the conceptual confusion that typifies most radiological studies of self.

A. Finding the Self: From its inception, radiological studies of the self have been attempting to locate *it* in the brain (e.g., Craik et al., 1999; Kelley et al., 2002). Originally broad in their objectives (i.e., finding *the* self; for reviews see Gillihan & Farah, 2005 and Northoff et al., 2006), several recent efforts have narrowed their search parameters, focusing on contextualized aspects of self (e.g., the depressive self; e.g., Lemogne et al., 2009).

But is the question "where in the brain is the self?" the right question to ask of nature? Judging from the proliferation of scanning studies that have attempted such localization (for review, see Northoff et al., 2006), researchers apparently believe that is. However, as I hope I have made clear, there is no self– either neurally or conceptually -- to be localized. Rather, there are at least two aspects of self – the objective and subjective -- and within the former, at least six functionally independent sub-systems have been identified (e.g., S. Klein, 2010, 2012a; Martinelli et al., 2013; Picard et al., 2013). Thus, the self is not a *thing* capable of localization: Rather it consists in a set of contents, qualities, functions and aspects, not all of which are amenable to scientific investigation (see the section "The Self as the Object of Scientific Inquiry" for discussion). Small wonder meta-analytic assessments of the success of localization via imaging technologies have thus far been disappointing (e.g., Klein et al., 2010; Northoff et al., 2006; Ruby & Legrand, 2007).

Considered in a more positive light, the studies seem to be telling us is that the self is not a single thing amenable to neural localization. If investigators had a more sensitive appreciation of this multiplicity, their attempts to map construct to observation presumably would be more nuanced and circumscribed. Unfortunately, with very few exceptions (e.g., Gillihan & Farah, 2005), discourse found in articles attempting to locate "the" self has been concerned primarily with methodological and anatomical issues. Theoretical considerations relevant to the conceptual target of investigation, in contrast, are typically accorded cursory treatment (you can confirm this for yourself by taking a look at *any* of the articles I have cited as examples of the radiological analysis of "self-in-the-brain" or "self-reflection". Non-cited papers on these – and other – topics involving radiological investigation of the self evidence a similar imbalance in attention given to method and theory).

This is a serious problem for the successful conduct of a scientific investigation of nature. As I hope I have shown, the specificity with which the constructs we submit to scanning capture the essence of what they are intended to describe carries at least as much weight as do methodological questions concerning the specificity with which radiological techniques reveal the neural activity assumed to be correlated with the construct of interest (for discussions of the latter point, see Miller et al,. 2002; Uttal, 2001; Vul, Harris, Winkielman, & Pashler, 2009).

B. The Neural Localization of Self-Reflection: Radiological studies of cortical involvement in the act of self-reflection also have been the subject of considerable attention (e.g., Herwig, Kaffenberger, Schell, Jancke, & Bruhl, 2012; Jenkins & Mitchell, 2011; Johnson, Nolan-Hoeksema, Mitchell, & Levin, 2009; Johnson et al., 2006). And, as was the case with attempts to "find the self", studies devoted to uncovering the neural underpinnings of self-reflection often contain little, if any, discussion of conceptual and philosophical issues associated with their presumed target. Rather, almost all attention is focused on questions pertaining to methodology and neuroanatomy.

This is particularly troubling given the fact that serious conceptual issues argue against either the logical or empirical possibility of separating the subjective aspects of the self from its more objective aspects. To the extent that self-reflection absent an object of reflection is a logical impossibility (i.e., the intentionality problem), there is no principled way of identifying whether enhanced neural activity during self-reflection maps neural activity associated with (a) the object(s) of subjectivity, (b) subjectivity per se, (c) both, or (d) neither.

In addition, ontological issues raised earlier in this paper argue against the possibility of

subjectivity being taken as an object of scientific inquiry (at least if that inquiry requires the current scientific "gold standards" of objectification and quantification).The first-person aspect of self forfeits its subjectivity when treated as an object of awareness. In effect, the attempt to "picture" subjectivity has the consequence of the stripping the image of its connection with the aspect of self (i.e., subjectivity) it was designed to address. Assuming that some neural structures are likely to be differentially active during self-reflection, what do those differences tell us? Some strong philosophical arguments are needed before we even can begin to formulate an answer. And such arguments are hard to find (see Legrand & Ruby, 2009, for one particularly salient exception).

7. FINAL THOUGHTS

Either you know what you are looking for, and then there is no problem, or you do not know what you are looking for, and then you cannot expect to find it. Without a nuanced, conceptually sophisticated analysis to guide our empiricism, asking the "right" questions of nature becomes a daunting task. And, absent the "right" questions, the answers nature provides are unlikely to instill a rationally justifiable sense of confidence that its joints have conceptually been severed.

The issues raised in this paper have been addressed to interpretive ambiguities characterizing the relation between image and theory in the social neurosciences. However, the same considerations apply with roughly equal force to other methodologies: Any attempt to map physical observation to abstract theory requires that attention be given in at least equal measure to both sides of the mapping. Indeed, I suggest that the balance should be shifted toward theory when the observational side of the mapping involves radiological measures. There are several reasons for this. Most important, perhaps, is the fact that most social and cognitive neuroscientists view psycho-neural identity theory (e.g., Place, 1956) and its variants as metaphysical dogma (for review and discussion, see Churchland, 1986; Kirk, 2003; Klein, in press). Conjoined with the reductionist approach of Western science (e.g., Bickle, 2003), identity theory has considerable potential to (mis)lead researcher and reader alike to assume an equivalence between construct and observation: i.e., that the neural structures activated during radiological analysis are in some (unspecified) way coextensive with the construct under consideration. With regard to the self, it appears that the temptation is strong to assume that when looking at images of the brain we are looking at *the* self. But, it is one thing to argue that neural activity is correlated with mental activity, and quite another to presume the two are coextensive.

In my experience, experimental treatments of the self (as well as many other constructs of interest in the social sciences) too often rely on folk-psychological familiarity (i.e., Russell's "knowledge by acquaintance") to compensate for the absence of well-specified, theoretical treatments of the construct under consideration (Klein, 2013). If this assessment is accurate (and I think that it is), things need to change. Absent clear theoretical guidance, the relation between observation and construct is difficult, if not impossible, to state with any conceptual warrant. We need to know not only what our experimental operations entail but also what the underlying constructs they are assumed to elucidate consist in. And nowhere is that more important than in neuroimaging studies, where the image runs the serious risk of being taken for the construct (e.g., Dumit, 2004; Klein, 2010-b; Uttal, 2001).[6]

So, what can be done? One possible remedial step would be to require researchers who study the self to step out of their comfort zone. Rather than assume their readers are familiar with what it means to be a self, authors should make explicit the manner in which they use the term -- that is, they need to provide explicit, theory-based treatments of how the term "self" is being conceptualized in the research under consideration. Otherwise, the term runs the very real risk of being reduced to little more than a place holder in a two-term relation (e.g., self-hyphen-topic of interest) or to a contextually limited meaning (e.g., cultural self, regulative self, narrative self) — its use predicated

[6] The critiques raised in this paper are quite specific and should not be taken as a global indictment of the use of radiological measures in social neuroscience. Insights obtained from a neural level of analysis can, and should, inform research at the psychological level (e.g., Polster, Nadel, & Schacter, 1991). For example, scanning studies that offered convergent evidence for the functional independence of the proposed constituents of a model of the self (e.g., Klein, 2012a) are a welcome contribution to existing theory and an appropriate use of radiological technology (e.g., Klein, 2010a). The work of Martinelli and colleagues (e.g., Martinelli et al., 2013) is exemplary in this regard.

on the researcher's conviction that the reader, by virtue of his or her personal acquaintance and social convention, will not worry about the self's function, role, or possible reality.

Philosophy, as I see it, also plays a crucial role as we try to come to grips with the issues I have raised. In our haste to stake out a claim as an independent, empirically-based, science some 150 years ago (by 1870 the borders between philosophy and psychology already were beginning to be clearly drawn), psychology has strayed too far from its philosophical roots in Departments of Natural Science/Philosophy. It is time we reclaimed some of the ground lost during the birthing process.

While it would be unwise to accuse philosophy of having resolved many of the questions with which it grapples, it certainly is the case that philosophers can help psychologists to (a) see what questions need to be asked, (b) identify when questions are well-formed and internally consistent, and (c) determine whether interpretation of an outcome is plausible with regard to its logical coherence as well as its empirical constraints (for discussion, see Craver, 2007).

On the flip side, there clearly is a need for philosophers to pay greater attention to findings from psychology. Careful consideration of psychological empiricism offers an important means by which philosophers can constrain theories based largely on notions of conceivability and logical coherence. Psychological findings can provide grounding for philosophers trying to transition conceivability to empirical possibility (e.g., Wilkes, 1988).

As it currently stands, philosophy has the advantage in the endeavor to close the interdisciplinary gap: A number of philosophers already have taken steps to naturalize their subject matter by drawing on findings from the psychological sciences (e.g., Flanagan, 2002; Klein & Nichols, 2012; Knobe & Nichols, 2008). It is time for psychology to reciprocate — to take advantage of the substantial resources philosophy has to offer. We need to work in earnest to close the gap we created between these academic siblings, whose separation was based more on personal, pragmatic and political reasons than pedagogic concerns.

8. REFERENCES

Abraham, A. (2013). The world according to me: Personal experience and the medial Prefrontal cortex. *Frontiers in Human Neuroscience, 7*. doi: 10.3389/fnhum.2013.00341

Albahari, M. (2006). *Analytical Buddhism: The two-tiered illusion of self*. Houndsmills, UK: Palgrave Macmillan.

Allport, G. W. (1961). *Patterns and growth in personality*. New York, NY: Holt, Rinehart & Winston.

Baillie, J. (1993). *Problems in personal identity.* New York, NY: Paragon House.

Beauregard, M., Lévesque, J., & Bourgouin, P. (2001). Neural correlates of conscious self-regulation of emotion. *The Journal of Neuroscience, 21,* 6993-7000.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Boston, MA: Kluwer Academic Publishers.

Brentano, F. (1995). *Descriptive psychology*. London, UK: Routledge.

Bruner, J. (2002). *Making stories: Law, literature, life*. New York, NY: Farrar, Straus and Giroux.

Brunswik, E. (1947/1956). *Perception and the representative design of psychological experiments.* Berkeley,CA: University of California Press.

Calkins, M. W. (1915). The self in scientific psychology. *Psychological Bulletin, 12*, 495-524.

Carnap, R. (1936). Testability and meaning. *Philosophy of science, 3*, 420-468.

Chadha, M. (2013). The self in early Nyaya: A minimal conclusion. *Asian Philosophy, 23*, 24-42.

Chen, P-H. A., Wagner, D. D., Kelley, W. M., Powers, K. E., & Heatherton, T. F. (2013). Medial prefrontal cortex differentiates self from mother in Chinese: Evidence from self-motivated immigrants. *Culture and Brain, 1*, 3-15.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain.* Cambridge, MA: MIT Press.

Conway, M. A. (2005). Memory and the Self. *Journal of Memory and Language, 53*, 594-628.

Craik, F. I. M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A PET investigation of self-referential information. *Psychological Science, 10*, 26-34.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford, UK: Oxford University press.

Coltheart, M. (2006). What has functional neuroimaging told us about the mind so far? *Cortex, 42,* 323-331.

Dainton, B. (2008). *The phenomenal self*. New York, NY: Oxford University Press.

Day, J. M. (1994). *Plato's Meno in focus*. London, UK: Routledge.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown and Company.

Denny, B., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self and other judgments reveals a spatial gradient for mentializing in medial prefrontal cortex. *Journal of Cognitive Neuroscience, 24*, 1742-1752.

Dumit, J. (2004). *Picturing personhood: Brain scans and biomedical identity*. Princeton, NJ: Princeton University Press.

Eakin, J.P. (2008). *How our lives become stories*. Ithaca, NY: Cornell University Press.

Earle, W. (1955). *Objectivity: An essay on phenomenological ontology*. New York, NY: The Noonday Press.

Earle, W. E. (1972). *The autobiographical consciousness*. Chicago, IL: Quadrangle Books.

Flanagan, O. (2002). *The problem of the soul*. New York, NY: Basic Books.

Fodor, J. A. (1968). *Psychological explanation: An introduction to the philosophy of psychology*. New York, NY: Random House.

Forman, R. K. C. (1990). *The problem of pure consciousness.* Oxford, UK: Oxford University Press.

Freeman, M. (1991). *Rewriting the self: History, memory, narrative.* London, UK: Routledge.

Freud, S. (1949). *The Ego and the Id*. London, UK: The Hogarth Press Ltd.

Gallagher, S., & Shear, J. (1999). *Models of the self.* Thorverton, UK: Imprint Academic.

Ganeri, J. (2012). *The self: Naturalism, consciousness, & the first-person stance*. Oxford, UK: Oxford University Press.

Gergen, K. J. (1971). *The concept of self*. New York, NY: Holt, Rinehart, & Winston, Inc.

Giles, J. (1997). *No self to be found*. Lanham, MD: University Press of America.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*, 76-97.

Godfrey-Smith, P. (2003). *Theory and reality*. Chicago, IL: University of Chicago Press.

Greenwald, A. G. (1981). Self and memory. In G. H. Bower (Eds.), *The psychology of learning and motivation. Vol 15* (pp. 201-236). New York, NY: Academic Press.

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences, 98*, 4259-4264.

Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science.* Cambridge, UK: Cambridge University Press.

Heatherton, T. F. (2011). Neuroscience of self and self-regulation. *Annual Review of Psychology, 62*, 363-390.

Heatherton, T. F., Macrae, C.N., & Kelley, W.M. (2004). What the social brain sciences can tell us about the self. *Psychological Science, 13*, 190-193.

Heisenberg, W. (1999). *Physics and philosophy*. Amherst, NY: Prometheus Books.

Hempel, C. (1965). *Aspects of scientific explanation.* New York, NY: Free Press.

Herwig, U., Kaffenberger, T., Schell, C., Jancke, L., & Bruhl, A. B. (2012). Neural activity associated with self-reflection. *BMC Neuroscience, 13*. Retrieved from: http//www. biomedcentral.com/1471-2202/13/52.

Hood, B. (2012). *The self illusion: How the social brain creates identity*. New York, NY: Oxford University Press.

Hume, D. (1739-1740/1978). *A treatise of human nature*. Oxford, UK: Oxford University Press.

Husserl, E. (1964). *The phenomenology of internal time-consciousness.* Bloomington, IN: Indiana University Press.

Ismael, J. T. (2007). *The situated self*. Oxford, UK: Oxford University Press.

James, W. (1890). *Principles of psychology* (Vol.1). New York, NY: Henry Holt and Company.

Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of

self-reflection. *Social Neuroscience, 6*, 211-218.

Johnstone, H. W. (1970). *The problem of the self.* University Park, PA: The Pennsylvania State University Press.

Johnson, M. K., Nolen-Hoeksema, S., Mitchell, K. J., & Levin, Y. (2009). Medial cortex activity, self-reflection and depression. *Social, Cognitive and Affective Neuroscience, 4,* 313-327.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Touryan, S. R., Greene, E. J., & Nolen-Hoeksema, S. (2006). Dissociating medial frontal and posterior cingulate activity during self-reflection. *Social, Cognitive and Affective Neuroscience, 1,* 56-64.

Jopling, D. A. (2000). *Self-knowledge and the self.* New York, NY: Routledge.

Kant, I. (1998). *The Cambridge edition of the works of Immanuel Kant; Critique of pure reason*. (P. Guyer & A. W. Wood, Trans.). New York, NY: Cambridge University Press. (Original work published 1781)

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14,* 785-794.

Kihlstrom, J. F., Cantor, N., Albright, J. S., Chew, B. R., Klein, S. B., & Niedenthal, P. M. (1988). Information processing and the study of the self. In L. Berkowitz, (Ed.), *Advances in experimental social psychology, Vol. 21: Social psychological studies of the self: Perspectives and programs* (pp. 145-178). San Diego, CA: Academic Press.

Kihlstrom, J. F., & Klein, S. B. (1994). The self as a knowledge system. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition. Vol 1. Basic processes* (pp. 153-208). Hillsdale, NJ: Erlbaum.

Kihlstrom, J. F., & Klein, S. B. (1997). Self-knowledge and self-awareness. In J. G. Snodgrass & R. L. Thompson, (Eds.), *The self across psychology: Self-awareness, self-recognition, and the self-concept. Vol 818* (pp. 5-17). New York, NY: Annals of the New York Academy of Sciences.

Kim, K., & Johnson, M. K. (2010). Extended self: Medial prefrontal activity during transient association of self and objects. *Social, Cognitive and Affective Neuroscience, 7,* 199-207.

King, N. (2000). *Memory, narrative identity: Remembering the self*. Edinburgh, UK: Edinburgh University Press Ltd.

Kirk, R. (2003). *Mind and body.* Montreal, Ca: Mcgill-Queen's University Press.

Klee, R. (1997). *Introduction to the philosophy of science: Cutting nature at its seams*. New York, NY: Oxford University Press.

Klein, C. (2010a). Philosophical issues in neuroimaging. *Philosophy Compass, 5*, 186-198.

Klein, C. (2010b). Images Are Not the Evidence in Neuroimaging. *British Journal for the Philosophy of Science ,61*, 265-278.

Klein, S. B. (2001). A self to remember: A cognitive neuropsychological perspective on how self creates memory and memory creates self. In C. Sedikides & M. B. Brewer (Eds.), *Individual self, relational self, and collective self* (pp. 25-46). Philadelphia, PA: Psychology Press.

Klein, S. B. (2004). The cognitive neuroscience of knowing one's self. In M. A. Gazzaniga (Ed.), *The Cognitive Neurosciences III* (pp. 1007-1089). Cambridge, MA: MIT Press.

Klein, S. B. (2010). The self: As a construct in psychology and neuropsychological evidence for its multiplicity. *WIREs Cognitive Science, 1*, 172-183.

Klein, S. B. (2012a). The self and its brain. *Social Cognition, 30*, 474-516.

Klein, S. B. (2012b). The Sense of Diachronic Personal Identity. *Phenomenology and the Cognitive Sciences*. Retrieved from http://link.springer.com/article/10.1007%2Fs11097-012-9285-8?LI=true#page-1

Klein, S. B. (2013). *What can recent replication failures tell us about the theoretical Commitments of Psychology?* MS submitted to *Theory & Psychology*.

Klein, S. B. (in press). *The two selves: Their metaphysical commitments and functional independence*. New York, NY: Oxford University Press.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2001). Priming exceptions: A test of the scope hypothesis in naturalistic trait judgments. *Social Cognition, 19*, 443-468.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review, 109*, 306-329.

Klein, S. B., & Gangi, C. E. (2010). The multiplicity of self: Neuropsychological evidence and its implications for the self as a construct in psychological research. The Year in Cognitive Neuroscience 2010: *Annals of the New York Academy of Sciences, 1191*, 1-15.

Klein, S. B., & Kihlstrom, J. F. (1998). On bridging the gap between social-personality psychology and neuropsychology. *Personality and Social Psychology Review, 2*, 228-242.

Klein S. B., & Lax, M. L. (2010). The unanticipated resilience of trait self-knowledge in the face of neural damage. *Memory, 18*, 918-948.

Klein, S. B., Lax, M. L, & Gangi, C. E. (2010). A call for an inclusive approach to the social cognitive neurosciences. *Social Cognition, 28*, 747-755.

Klein, S. B., & Loftus, J. (1993). The mental representation of trait and autobiographical knowledge about the self. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition. Vol. 5* (pp. 1-49). Hillsdale, NJ: Erlbaum.

Klein, S. B., Loftus, J., Trafton, R. G., & Fuhrman, R. W. (1992). The use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology, 63*, 739-753.

Klein, S. B., & Nelson, C. R. (2014). The effects of self-reference on memory: A conceptual and methodological review of inferences warranted by the self-reference effect. In T. J. Perfect & D. S. Lindsay (Eds.), *The Sage handbook of applied memory* (pp. 256-272). Los Angeles, CA: Sage Publications Ltd.

Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind, 121,* 677-702.

Klein, S. B., Robertson, T. E., Gangi, C. E., & Loftus, J. (2008). The functional independence of trait self-knowledge: Commentary on Sakaki (2007). *Memory, 16*, 556-565.

Klein, S. B., Rozendale, K., & Cosmides, L. (2002). A social-cognitive neuroscience analysis of the self. *Social Cognition, 20*, 105-135.

Knobe, J., & Nichols, S. (2008). *Experimental philosophy.* Oxford, UK: Oxford University Press.

Ladyman, J. (2002). *Understanding philosophy of science*. New York, NY: Routledge.

Leahy, R. L. (1985). *The development of the self.* Orlando, FL: Academic Press, Inc.

Leary, M. R., & J. P. Tangney (2012). *The Handbook of Self and Identity.* New York, NY: Guilford Press.

Lecky, P. (1945). *Self-consistency: A theory of personality*. New York, NY: Island Press.

Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review, 116*, 252-282.

Lemogne, C., le Bastard, G., Mayberg, H., Volle, E., Bergouigan, L., Lehericy, S., … Fossati, P. (2009). In search of the depressive self: Extended medial prefrontal network during self-referential processing in major depression. *Social, Cognitive, and Affective Neuroscience, 4,* 305-312.

Leshikar, E. D., & Duarte, A. (2013). Medial prefrontal cortex supports source memory for self-referenced material in young and older adults. *Cognitive, Affective and Behavioral Neuroscience.* doi: 10.3758/s13415-013-0198-y

Lewis, H. D. (1982). *The elusive self*. Philadelphia, PA: Westminster Press.

Lund, D. H. (2005). *The conscious self*. Amherst, NY: Humanity Books.

Ma, N., Baetens, K., Vanderkerckhove, M., Kestemont, J., Fias, W., & Van Overwalle, F. (2013). Traits are represented in the medial prefrontal cortex: An fMRI adaptation study. *Social, Cognitive, and Affective Neuroscience.* doi: 0.093/scan/nst098

Macrae, C. N., Morgan, J. M., Heatheron, T. F., Banfiled, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex, 14*, 647-654.

Margenau, H. (1950). *The nature of physical reality.* New York, NY: McGraw Hill.

Martin, C. B. (2008). *The mind in nature*. Oxford, UK: Oxford University Press.

Martinelli, P., Sperduti, M., & Piolino, P. (2013). Neural substrates of the self-memory system: New insights from a meta-analysis. *Human Brain Mapping, 34*, 1515-1529.

McGinn, C. (1991). *The problem of consciousness: Essays toward a resolution*. Oxford, UK: Blackwell Publishers.

Meixner, U. (2008). The *reductio* of reductive and non-reductive materialism – and a new start. In A. Antonietti, A. Corradini, & E. J. Lowe (Eds.), *Psycho-physical dualism: An*

interdisciplinary approach (pp. 143-166). Boulder, CO: Rowman & Littlefield Publishers, Inc.

Metzinger, T. (2009). *The ego tunnel: The science of mind and the myth of the self*. New York, NY: Basic Books.

Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., …Gazzaniga, M. S. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience, 14*, 1200-1214.

Moran, J. A., Kelley, W. M., & Heatherton, T. F. (2013). What can the organization of the brain's default mode network tell us about self-knowledge? *Frontiers in Human Neuroscience, 7.* doi: 10.3389/fuhum.2013.00391

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*, 435-450.

Neisser, U. (1988). Five kinds of self–knowledge. *Philosophical Psychology, 1,* 35–39.

Neuhouser, F. (1990). *Fichte's theory of subjectivity*. New York, NY: Cambridge University Press.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). San Francisco, CA: Academic Press.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain —a meta-analysis of imaging studies on the self. *NeuroImage, 31*, 440-457

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Kihlstrom, J. F., D'Esposito, M & Gabrieli, J. D. E. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage, 28,* 797-814.

Olson, E. T. (1997). *The human animal: Personal identity without psychology*. New York, NY: Oxford University Press.

Quine, W. V .O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*, 20-43.

Parker, D. B. (2007). *The self in moral space: Life narrative and the good.* Ithaca, NY: Cornell University Press.

Pessoa, L., Thompson. E., & Noë, A. (1998). Filling-in is for finding out. *Behavioral and Brain Sciences, 21*, 781-796.

Picard, L., Mayor-Dubois, C., Maeder, P., Kalenzaga, S., Abram, M., Duval, C., … Piolino, P. (2013). Functional independence within the self-memory system: Insight from two cases of developmental amnesia. *Cortex , 49*, 1463–1481.

Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology, 47,* 44-50.

Polster, M. R., Nadel, L., & Schacter, D. L. (1991). Cognitive neuroscience analysis of memory: A historical perspective. *Journal of Cognitive Neuroscience, 3*, 95-116.

Popper, K. (1963/2004). *Conjectures and refutations: the growth of scientific knowledge*. London, UK: Routledge.

Prebble, S., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin.*

Rescher, N. (1997). *Objectivity: Obligations of impersonal reason*. Notre Dame, IN: University of Notre Dame Press.

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences, 16,* 550–558 .

Ruby, P., & Legrand, D. (2007). Neuroimaging the self? In Y. Rossetti, P. Haggard and M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (22nd Attention & Performance Meeting, pp. 293-318). New York, NY: Oxford University Press.

Russell, B. (1912/1999). *The problems of philosophy.* Mineola, NY: Dover Publications.

Samsonovich, A. V., & Nadel, L. (2005). Fundamental principles and mechanisms of the conscious self. *Cortex, 41*, 669-689.

Savoy, R. (2005). Experimental design in brain activation MRI: Cautionary tales. *Brain Research Bulletin, 67*, 361-367.

Schwerin, A. (2012). *Hume's labyrinth: A search for the self*. Newcastle upon Tyne, UK: Cambridge Scholar's Publishing.

Sedikides, C., & Brewer, M. B. (2001). *Individual self, relational self, collective self.* Philadelphia, PA: Psychology Press.

Siderits, M. (2003). *Personal identity and Buddhist philosophy*. Hampshire, UK: Ashgate Publishing Limited.

Siderits, M., Thompson, E., & Zahavi, D. (2011). *Self, no self: Perspectives from analytical*

phenomenological and Indian traditions. Oxford, UK: Oxford University Press.

Sorabji, R. (2006). *Self: Ancient and modern Self-awareness and alterity: A phenomenological investigation insights about individuality, life, and death.* Chicago, IL: The University of Chicago Press.

Stern, D. (1985). *The interpersonal world of the infant.* New York, NY: Basic Books.

Strawson, G. (2009). *Selves: An essay in revisionary metaphysics*. New York, NY: Oxford University Press.

Strawson, G. (2011). *The evident connection*. Oxford, UK: Oxford University Press.

Stuss, D. T. (1991). Self, awareness and the frontal lobes: A neuropsychological perspective. In J. Strauss & G. R. Goethals (Eds.), *The self; Interdisciplinary approaches* (pp. 255- 278). New York, NY: Springer-Verlag.

Swinburne, R. (1997). *The evolution of the soul*. Oxford, UK: Claredon Press.

Swinburne, R. (2013). *Mind, brain, and free will*. Oxford, UK: Oxford University Press.

Symonds, P. M. (1951). *The ego and the self*. New York, NY: Appleton-Century-Crofts, Inc.

Textor, M. (2013). Brentano on the dual relation of the mental. *Phenomenology and the Cognitive Sciences, 12*, 465-483.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York, NY: John Wiley & Sons, Inc.

Trusted, J. (1979). *The logic of scientific inference.* London, UK: The Macmillan Press Ltd.

Uttal, W. R. (2001). *The new phrenology*. Cambridge, MA: MIT Press.

Vierkant, T. (2003). *Is the self real?* London, UK: Transaction Publishers.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*, 274-290.

Williams, B. (1973). *Problems of the self*. New York, NY: Cambridge University Press.

Wilkes, M. (1988). *Real people: Personal identity without thought experiments*. Oxford, UK: Claredon Press.

Wittgenstein, L. (1953/2009). *Philosophical investigations* (Anscombe, G. E. M., Hacker, P. M. S., & Schultze, J. Trans.). West Sussex, UK: Wiley-Blackwell. (Original work published 1953)

Yao, Z. (2005). *The Buddhist theory of self-cognition.* New York, NY: Routledge.

Zahavi, D. (1999). *Self-awareness and alterity: A phenomenological investigation.* Evanston, IL: Northwestern University Press.

Zahavi, D. (2005). *Subjectivity of selfhood: Investigating the first-person Perspective*. Cambridge, MA: The MIT Press.