Self-Consciousness and the Self-as-Subject

by

Ted Lougheed

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Cognitive Science

Carleton University

Ottawa, Ontario

Abstract

The overarching purpose of this project is to expand upon our understanding of the nature of self-consciousness. Specifically, I investigate our sense of self as a *subject of experience*, as one and the same experiencer extended over time. Conceptual inquiry and cognitive psychology studies are the primary means of investigation. I seek to fulfill two primary goals. First, I make the case for the distinction between consciousness of the self-as-subject (SAS) and self-as-object (SAO). I lay forth the conceptual foundations for this distinction and discuss the properties of SAS consciousness. I also review empirical literature that reveals, intentionally or not, the importance of the distinction.

Second, after reviewing the present state of existing studies pertaining to SAS consciousness, I present a complete study and a preliminary study that I undertook to identify and investigate the mechanism responsible for the phenomenon. I examine the relationship between SAS consciousness and other cognitive faculties, specifically episodic memory and attention. I claim that episodic memory is dependent upon SAS consciousness, and that attention must be paid to self for episodic memory to form. In the first study, I examine the effects of distraction on SAS consciousness in adult participants. In the second, preliminary study, I turn my attention to children, investigating how SAS consciousness develops in 3-5 year-olds, with respect to other aspects of self-consciousness.

Acknowledgements

I would like to thank my supervisor, Andrew Brook, for his guidance and encouragement that began before I even joined the program. I am honored to have had the opportunity to work with him. I would also like to thank Deepthi Kamawar for all her help in learning the ways of empirical research and especially research with young children, with all the trials and tribulations that entails. I would also like to thank the women in CRDL, Corrie Vendetti, Andrea Astle, Gal Podjarny and Kim Connolly for letting me share their space and helping me out through all stages of research. I would like to thank John Logan for his guidance during the design and testing of my studies. His feedback has been invaluable in helping me navigate the waters of empirical research. Thank you to all of my friends and colleagues who volunteered their time to help me pilot my studies.

Thank you to the Dean of Graduate Studies Entrance Scholarship, Carleton University Graduate Scholarship, Social Sciences and Humanities Research Council, Ontario Graduate Scholarship, Agnes M. Ireland Memorial Fund, David and Rachel Epstein Foundation, Dr. Thomas Betz Memorial Fund, Hamlin Graduate Fellowship, and the Cognitive Science Graduate Scholarship (now the Andrew Brook Graduate Scholarship) for financial support.

I would like to thank my wife for her love, support, and patience throughout this journey, and my parents for supporting me in so many ways and encouraging me to follow my aspirations and never settle for less. I would also like to thank my sisters and their families for their support and helping me recruit children for one of my studies.

This work is dedicated to the memory of my grandmother, Mary, who passed away during the project.

Table of Contents

List of Tables

List of Illustrations

List of Appendices

Chapter I – Conceptual and Empirical Background

Introduction

Research Goals

The overarching purpose of this project is to expand upon our understanding of the nature of self-consciousness. Specifically, I investigate our sense of self as a *subject of experience*, as one and the same experiencer extended over time. Conceptual inquiry and cognitive psychology studies are the primary means of investigation. I seek to fulfill two primary goals. In the first half, I will make the case for the distinction between consciousness of the self-as-subject (SAS) and self-as-object (SAO), a distinction on which I will elaborate shortly. I put forth the conceptual foundations for this distinction and discuss the properties of SAS consciousness. I will also review empirical literature that reveals, intentionally or not, the importance of the distinction.

In the second half, I review selected literature on the indicators of SAS consciousness. In doing so, I will also examine the relationship between SAS consciousness and other cognitive faculties, specifically episodic memory and attention. I claim that episodic memory is dependent upon SAS consciousness, and that attention must be paid to self for episodic memories to form. I describe two studies that I undertook to identify and investigate the mechanism responsible for SAS consciousness. In the first study, I examine the effects of distraction on SAS consciousness in adult participants. In the second, preliminary study, I turn my attention to children, investigating how SAS consciousness develops in 3-5 year-olds, with respect to other aspects of self-consciousness.

Terminology

Before I can begin this investigation, I need to clarify how I am using the term "self-consciousness". The study of self-consciousness is plagued with terminological confusion, inherited from the study of consciousness in general. This confusion stems from a general lack of agreement concerning the meaning of the term "consciousness," with the result that many theorists fail to refer to the same phenomenon when they use the term. As Seager (2007) notes, it is unclear if there is any common ground among all the different uses of the term. We may take consciousness to mean a special kind of activity, or choose instead to focus on the phenomenal character of experience as the central feature of consciousness. On the other hand, we may take consciousness to be more than just the raw "feels" of subjective experience, and include the presence of evaluative states as a critical feature of consciousness (Seager, 2007, pp. 9-10).

As a starting point, we can understand someone being conscious *of* something as the thing *being like something* to the person (Nagel, 1974; Brook & Raymont, forthcoming). To be conscious, then, is to be capable of having things such as external objects and mental states appear to me as "like something." Unfortunately, this is the most that I can say about the term 'consciousness' that theorists can agree upon.

Some theorists distinguish different types of consciousness. Brook and Raymont (forthcoming), for instance, make a distinction between self-consciousness, consciousness of world, and consciousness of one's own states. They allow for the possibility that beings such as nonhuman animals could be conscious of the world around them but unconscious of their mental states. A hungry animal, for example, may be conscious of food in the surrounding environment, or conscious of its own hunger, yet fail to be conscious of itself *as the thing that is hungry*.

On the other side are those theorists who limit their concept of consciousness to a single type. Said theorists would not, for instance, deny the existence of consciousness of world, but would deny that such a phenomenon deserves to be called ‘consciousness’ at all. Drew McDermott, for example, holds that consciousness is merely introspection about one's own states, which he also equates with self-consciousness (2007, p. 138). On this view, I cannot be conscious of something without being conscious of being conscious of that thing. McDermott would suppose that, for example, consciousness of things in the world is really consciousness of mental states *about* one’s perceptions, rather than consciousness of the percepts themselves. Similarly, consciousness of oneself as the subject of one's experiences is simply consciousness of a particular mental state, namely being in a self-conscious state.

McDermott's theory can be considered a HOT (higher-order thought) theory of consciousness, because it posits the need for meta-representations, that is, representations of representations. For McDermott, what Brook would call consciousness of world is not considered consciousness at all, but a kind of “raw” representation - it is not “like” anything to the organism unless that representation is, in turn, represented to the organism. The problem with HOT theories like McDermott's is that they can lead to a kind of infinite regress. If a raw representation does not constitute consciousness, we need to ask what is special about a representation *of* some lower-order representation, that is, why it is that a representation of a further representation is consciousness, but a single level of representation fails to be. So long as we are not given a reason, we could continue to posit additional levels of representation, such as a representation of a representation of a representation.

The general lack of agreement concerning the meaning of ‘consciousness’ and ‘self-consciousness’ is important to bear in mind when examining empirical work on consciousness.

To work around this problem, I will identify the similarities and differences between my definition of consciousness and that of the researcher in question. This task is especially difficult since some researchers are resistant to precise definitions. Crick and Koch, for instance, argue that "[u]ntil the problem is understood much better, any attempt at a formal definition is likely to be either misleading or overly restrictive, or both." (2003, p. 35) Crick and Koch go so far as to urge us to set aside conceptual questions about consciousness and focus on the science.

To advance the field beyond speculation, some psychologists have attempted to provide operational definitions of consciousness. Baars, for example, gives a working definition of conscious processes as events that can be reported and acted upon with verifiable accuracy, under optimal reporting conditions, and that are reported as conscious (2003, p. 1). This definition is intuitively plausible because of the common-sense notion that we must be conscious of something in order to report on it. The difficulty of identifying consciousness with reportability, however, is that it is unclear if the reverse relation holds, namely if we must be able to report on something in order to be conscious of it. As Brook and Raymont (forthcoming) note, many mental events are too short-lived to accurately report on. It is at least plausible to suppose that there is still something it is like to have them, however fleeting (forthcoming).

Following Brook and Raymont (forthcoming), I define consciousness as the appearance of external objects, mental states, and self to an organism, such that these are 'like something' to the organism. Self-consciousness, then, is the appearance of self such that it is like something to the organism. The current project seeks a balance between conceptual inquiry and empirical investigation, setting out from the beginning a set of criteria that phenomena must meet for inclusion in my study. Contrary to the opinion of Crick and Koch (2003), I maintain that there is still important conceptual work to be done, if only to avoid confusing what one theory means by

'consciousness' with that of another theory. I have chosen the above definition because it is sufficiently broad to capture what many researchers mean, yet narrow enough to distinguish the phenomenon from most of cognition.

What is Self? SAO and SAS

Issues surrounding the definition of consciousness aside, we must also be clear about what, exactly, we are conscious *of* when we talk about self-consciousness. I delineate two senses of the term 'self' which are closely related but important to distinguish—the 'self-as-subject' (SAS) and 'self-as-object' (SAO). The sense of SAS is the phenomenon of interest in this study, but first it is important to show how it contrasts with the sense of SAO.

Roughly, SAO can be defined as the set of features or properties—for example my bodily appearance or mental states and the like—that I identify with myself, or of which I take ownership. When I think of myself as SAO, I am thinking of a particular object in the world, namely a human being named Ted, with a certain height, weight, hair colour, and so forth. SAO consists of the properties I associate with myself. At any given moment I may hold in mind a small subset of these properties. I specify "at any given moment" because I cannot possibly hold all of the features that I associate with myself in mind at once. My sense of SAO changes over time as I come to attribute different features to myself. The features that I attributed to myself as a child are not the same features I attribute to myself now, so my sense of SAO may be different in almost every respect. Indeed, the features I attribute to myself can change daily, or from one moment to the next. I do not have a static conception of myself as an object that I call up whenever I think about myself; at one moment I may have a rich conception of myself in mind while at others, such as during a dream, I may have only a hazy concept of myself consisting of one or two significant features. These features may or may not accurately reflect actual

properties of the middle-sized object in the world I happen to embody. My physical appearance, political opinions, taste preferences, propensity to particular emotions, and other particular characteristics are all features that might compose one's sense of SAO at a given time.

The focus of this study is on the more distinctive sense of SAS. Briefly, the sense of SAS is the sense I have of myself as the *subject* of particular experiences, that is, the entity who has the experiences. The experiences that one has may or may not include the experience of oneself as an object, that is, SAO. What makes SAS distinctive is that it is part of the *structure,* rather than the *content* of the experience; it is not a particular experience *in addition to* the experience of oneself as an object, but the substrate that makes possible identification of particular features as belonging to oneself rather than another. Although people can be aware of SAO in ways that they are not normally aware of external objects (e.g., via proprioception), both consciousness of SAO and consciousness of the external world concern the content of experience. If SAO is comprised of particular features that we take ownership of, then SAS is what does the *owning*. In the discussions that follow, I will occasionally refer to the sense of SAS as the sense of ownership.

To have consciousness of SAS is to appear to oneself *as* oneself, to use Brook's phrasing (2001, p. 11). Consider a particular conception I have of myself at the time of writing. It is not enough to be aware of a constellation of specific features such as those listed above – I must also recognize that those features are *my* features, rather than those of somebody else. Suppose that I have been struck with severe retrograde amnesia, and fail to remember what I look like. Further suppose that I am strapped into an apparatus that restricts me from viewing any of my own body parts. If my captor were to show me a picture of myself, I would have no way of recognizing that the picture was a picture of me. Yet I still have an idea of me, namely the person viewing the

picture. Even if there is no information available about what I look like, what kind of person I am, or what my personal history might be like, I still have a sense of *me* as the person having experiences; this is the sense of SAS.

It is important to note at the outset that I am only committed to an epistemological distinction between SAO and SAS, rather than an ontological one. In other words, I am claiming that there are at least two distinct kinds of knowledge I can have of myself, yet not committing to there being two separate, corresponding psychological mechanisms. The distinction between conceiving of oneself as a subject or as an object is familiar within philosophical circles, and has been around at least since Kant (1781/1998, B135), but was explicitly introduced by Wittgenstein in *The Blue Book* (1934/1960) in the context of the use of the word "I". Wittgenstein argues that we use the word "I" in two ways, one to refer to self as an object in the world, and the other to refer to self as the subject of experiences. I will expand on this in the *Conceptual* section below.

The first goal of my project is to establish a conceptual and empirical basis for the SAS/SAO distinction. To this end, I claim that existing studies sufficiently demonstrate the necessity for such a distinction. Though the distinction is well-known to many philosophers of mind, it fails to enjoy the same familiarity in other cognitive disciplines. Most empirical study of the self and self-consciousness has thus far focused on what I have called SAO, often called the self-concept in psychological literature. Dorothée Legrand (2007) is one of the few psychologists to explicitly mention the distinction between SAO and SAS. Legrand defines SAS as the conception of myself as a subject of experience, contrasting it with SAO, which is the content of my experience that I conceive of as being part of self as opposed to part of the external world. She explains further that "[f]or *x* to be given as an object of experience is for *x* to be given as

differing from the subjective experience that takes it as an object. In short, subject and object of experience are necessarily different from each other: an object of experience is something that stands in opposition to the subject of experience." (p. 586) If *x* is a particular notion of the self, such as being a certain height or hair colour, this object of my experience cannot be equivalent to what is doing the experiencing. Neither Legrand nor I use "object" here in the ontological sense; by "object" we mean the contents of experience, whether or not they correspond to actual objects in the world. I will examine further empirical evidence for the SAS/SAO distinction in the *Empirical Support* section.

The second goal of this study is to identify and investigate behavioural indicators of SAS consciousness. I will use both conceptual and empirical methods to accomplish this task. In the first section of Chapter I, *Conceptual*, I argue that there are three main characteristics of SAS consciousness that set it apart from other mental phenomena. First, reference to SAS is indexical, meaning that it "points to" the subject of experience. Second, SAS is universally available—that is, available in any experience whatsoever—because it is independent of particular representations. Lastly, reference to SAS is non-attributive, meaning that we do not need to identify particular features in order to make reference to it, a characteristic that is closely connected to universal availability.

The empirical component of this project begins with an analysis and interpretation of existing behavioural studies. In the last section of Chapter I, *Empirical Support for SAS/SAO*, I provide a critical review of studies that make use of the SAS/SAO distinction, and those that would benefit from recognizing the distinction. My aim in that section is to establish the validity and importance of the distinction, as my research takes the distinction for granted.

In the first half of Chapter II, *Research on Indicators of SAS Consciousness*, I review empirical work on the nature of SAS consciousness, this time with a focus on those studies that provide evidence of the three distinctive characteristics I identify in the *Conceptual* section. I review work done on use of the first-person pronoun, the linguistic form of indexical reference to SAS. I also review research on episodic and autobiographical memory as it pertains to the availability of SAS in experience. Finally, I take a brief look at the work on attentional load and its bearing on SAS consciousness—I claim that people are deficient in SAS consciousness when they are distracted from self-related thought.

Following the review, I present my own set of behavioural studies on the development and nature of self-consciousness. One goal of this component of my project is to determine what behavioural conditions, if any, are unique to the phenomenon of self-consciousness. My research consists of novel behavioural studies of self-awareness in adults and young children (the latter between the ages of 3 and 5 years). Since I am using behavioural methods, I will be studying the mechanisms of self-consciousness indirectly.

Due to practical constraints, I cannot perform studies on the initial onset of self-consciousness in infancy. As I will show in the *Research on Indicators of SAS Consciousness* section in Chapter II, there is compelling evidence that children are SAS conscious by the middle of their second year, and I do not have access to a population in a suitable age range. Therefore, I must focus on a later stage of development. One of the aims of my preliminary developmental study is to investigate when SAO becomes integrated with SAS, in relation to developmental signposts such as verbal ability.

An example of the integration of SAS and SAO into a coherent sense of self is the stage where a child understands that the child in a photograph is *her*, that is, the body (SAO) belonging

to the subject of experience (SAS), only at an earlier time. Prior to such integration, the child can readily name her image in the photograph—as Robyn, for example—yet fail to understand that Robyn is *her*, because she cannot comprehend herself as an object extended into the past. Children may possess SAS consciousness from an early age, but they often demonstrate confusion about delayed self-imagery—photographs or videos—as late as age five (as demonstrated in Povinelli, 2001, which I review below). This suggests a deficit in their sense of self, but one specific to their understanding of themselves as objects extended into the past. If sense of self consisted simply of the sense of a particular object—the image of one's own body, say—it would be difficult to make sense of children's confusion over delayed self-imagery, as they have no trouble with present self-imagery. Fortunately, the SAS/SAO distinction allows us to make sense of the delayed self-image as a failure of integration between consciousness of SAS and SAO.

Through my investigation into children's ability to integrate consciousness of SAS and SAO, I hope to unravel more detail about the manner in which SAO is related to SAS, and how SAS relates to self-referential faculties. Briefly, I evaluate children's capacity for SAS consciousness and their ability to integrate SAO, by comparing their performance on a delayed self-imagery task with performance on measures of SAS consciousness, such as ability to refer to self indexically, and to recall episodic memories. As I will argue in my review of the relevant literature, episodic memory is dependent upon SAS consciousness, such that SAS consciousness must develop prior to the capacity for episodic memory.

The dependence of episodic memory on SAS consciousness is also of import to my first study. That study consists of research on adult participants (undergraduate psychology students). In this part, I seek to identify states of reduced SAS consciousness and compare those with self-

conscious states. I postulate that by engaging participants in attentionally demanding non-self-related tasks, I can distract them from self-reflective thoughts. Participants in this low self-reflection group are compared with a high self-reflection group who engaged in the same task, but with a constant reminder of self—the presence of self-images—encouraging self-reflective thought. If episodic memory is dependent of SAS consciousness, I expect that a reduction in SAS consciousness should reduce the amount of detail encoded in episodic memory, for any event experienced while SAS consciousness is reduced.

Conceptual

The conceptual component of this study seeks to accomplish two primary goals. The first goal is to address the confusion concerning the use of the terms "consciousness" and "self-consciousness." The aim is to settle on a clear definition of consciousness that avoids the circularity endemic to many such definitions. The second goal is to identify and elaborate on the key features of the mechanism behind SAS consciousness. The latter goal, as you may recall, is one of the primary goals of the entire project; here, I will explore the question to the fullest extent possible using conceptual tools alone. In the *Research on Indicators of SAS Consciousness* section in Chapter II, I will review what existing empirical work tells us about this mechanism; in the *Empirical Studies* section of that chapter, I describe a pair of studies that investigate my claims.

The Definition of Consciousness

Recall how I defined consciousness above, as how external objects, mental states, and self are "like something" to an organism. There are a number of things (some percepts and mental states, and most of the brain's processing) that fail to appear to us, even if they have some influence over the workings of the mind. I fail to be conscious, for example, of the inner

workings of many of my cognitive capacities, let alone more primary capacities such as the maintenance of my respiratory and pulmonary systems by the medulla. Much of what occurs in my brain fails to be presented to the conscious mind, that is, fails to appear to me.

We must be careful not to confuse the term ‘appearance’ with ‘perception.’ It is neither tantamount to nor restricted to perception. By ‘perception,’ I mean the processing and interpretation of stimuli. This excludes things like mental states that we introspect but are not directly or immediately related to stimuli. First of all, there may be some perceptions of which you fail to be conscious; that is, having a perception of something is insufficient for us to be conscious of it. Consider the case of motion-induced blindness. This is a simple optical illusion created by superimposing three stationary yellow spots over a rotating background of blue stripes. When you fixate on any particular part of the image for a few seconds, some or all of the yellow dots will seem to flash in and out of existence. Of course the dots are still there, and they are picked up by the retina and at least partially processed by the visual cortex (see Koch, 2004, pp. 14, 270-1). So while the dots are *perceived* on some level—they are processed to some extent by the visual cortex—they fail to *appear* to you.

Second, we are conscious of more than just perceptions—a mental state is a good example of something that you are conscious of but is not a perception. You can, for instance, be depressed without perceiving your depression, yet you are clearly conscious of being depressed, through introspection. More importantly for my purposes, as I shall argue below, you may be conscious of self independently of any particular perception. In sum, it is neither necessary nor sufficient to perceive something to be conscious of it, and consciousness need not be of perception—it can apply to other constructs, such as concepts or judgments. My use of the term

'appearance' rather than 'perception' is intended to capture what is unique about conscious states.

Self-as-Subject and Self-as-Object

SAS/SAO is the key distinction around which my entire project is centered—and one that I treat as a basic assumption in my empirical investigations. In this section and the next, I will show that there is ample evidence available for the distinction. Having done so, I will take the distinction for granted in Chapter II. The aim of my studies is not to make the case for the distinction, but to elaborate on the nature of SAS consciousness.

Under the notion of consciousness defined in the previous section, self-consciousness can be defined as the appearance of self, such that it is like something to the organism. This statement needs qualification, for the idea of self is also poorly defined. When I say I am conscious of myself, there are two possibilities of what I mean. When I look in the mirror, I am conscious of a particular object, namely myself, with particular features such as hair colour, height, body shape, and so forth. Turning inward, I may also reflect on other features of myself that fail to be outwardly visible, such as my current disposition or my musical preferences. I am also aware of being in a particular mental state. Taken together, the features I attribute to myself constitute what I call the 'self-as-object' (SAO). Recall from the *Introduction* that the sense of SAO is the sense of self, at any given moment, as a set of particular objects of experience. I am not claiming that at any given moment I have a complete and accurate description of myself, only that I readily identify features of my experience as either belonging to myself or to the world around me. As such, my experience of SAO is constantly changing as I become aware of, for example, different body parts, a feeling of hunger, or memories from my childhood. I may even mistakenly identify features of others as features of myself, or vice versa. I might happen

upon a mirror unexpectedly, for example, and momentarily mistake the image in the mirror for another person.

Cases where people have made errors in attributing body parts or other aspects of SAO to themselves or to others have been studied extensively in the empirical literature. A famous example is the rubber-hand illusion, as described in Botvinich and Cohen (1998). In this illusion, experimenters apply brush strokes to the participant's hand—hidden behind a screen—and to a rubber hand placed in a slightly offset location. When the brush strokes occur simultaneously, a participant will typically report feeling as if the rubber hand is her own. More recently, Petkova and Ehrsson (2008) investigated a similar phenomenon, extended to include the entire body, rather than a single body part, in the so-called "body-swap" illusion. Their study demonstrated that people will quite readily take ownership of other bodies when those bodies are "swapped" via a virtual-reality headset connected to a set of stereoscopic cameras. The effect works when the replacement bodies share fairly crude physical similarities with the participants' own—for example, participants will feel a sense of ownership over a mannequin (i.e., attribute the mannequin *to* self) or even a real person of a different sex, but not over a large green box.

Moving away from bodily attributions towards more abstract forms of attribution, Daprati, Franck, Georgieff, Proust, Pacherie, Dalery, and Jeannerod (1997) investigated the phenomenon of misattribution of actions to others in schizophrenics, with neurotypical participants as controls. In this experiment, a participant placed one hand out of sight and performed a simple hand or wrist movement. A video display relayed the image of the hand to the participant, only in some trials the image relayed was actually the hand of someone else, performing either the same movement or a different one. They found that schizophrenic patients, particularly those with delusions or hallucinations, would frequently misattribute the actions of

others to themselves. Surprisingly, Daprati *et al*. were able to elicit similar mistakes in neurotypical participants, albeit with much lower frequency. Jeannerod and Pacherie (2004) expands on the theoretical aspects of this work. I will have more to say about these cases in the *Empirical Support* section, below.

There is a second possibility as to what I mean when I say that I am conscious of myself, but it is harder to identify, and many have failed to notice it. Much of the existing empirical work done on self-consciousness has focused on SAO, often under the label of the self-concept. I am also, however, conscious of myself as the *subject* of my experiences, what I call the 'self-as-subject' (SAS). These experiences may, and often do, include features associated with myself (i.e., SAO). From moment to moment, the content of my experience can shift dramatically; in some cases, it bears almost nothing in common with a previous moment. Compare the experience of my being in a quiet room, reading a book, to skydiving from an airplane at 30,000 feet. My senses provide me with a different set of stimuli, and I am undoubtedly in a different mental state in one instance than the other. Yet there does seem to be something in common among these disparate experiences, namely that they are all experienced by *me*, and that I become conscious of these experiences merely by having them.

Unlike SAO, little empirical research has been done on SAS. One reason for this could be that there are few instances where awareness of SAS appears to break down. One important exception is the work done on the phenomenon of thought insertion, a relatively common symptom of schizophrenia. A largely theoretical treatment of this topic can be found in Stephens and Graham (2000). I will expand on this phenomenon in detail later in this section.

The idea that there are two senses of the term ‘self’ is by no means novel. As I explained in the introduction, the distinction between subject and object is familiar in philosophical circles. It dates at least as far back as Kant (1781/1998), who noted that:

> [H]ow the I that I think is to differ from the I that intuits itself… and yet be identical with the latter as the same *subject*, how therefore I can say that I as intelligence and thinking *subject* cognize my self as an *object* that is thought, insofar as I am also given to myself in intuition, only, like other phenomena, not as I am for the understanding but rather as I appear to myself, this is no more and no less difficult than how I can be an *object* for myself in general and indeed one of intuition and inner perceptions. (Sec. II, B155-6; my emphasis)

The point that Kant is making is tangential to the current discussion, but it demonstrates use of the distinction. More recently, we find mention of the distinction in Wittgenstein (1934/1960):

> [T]he idea that the real I lives in my body is connected with the peculiar grammar of the word “I”, and the misunderstandings this grammar is liable to give rise to. There are two different cases in the use of the word “I” (or “my”) which I might call “the use as object” and “the use as subject”. Examples of the first kind of use are these: “My arm is broken”, “I have grown six inches”, “I have a bump on my forehead”, “The wind blows my hair about”. Examples of the second kind are: “*I* see so-and-so”, “*I* hear so-and-so”, “*I* try to lift my arm”, “*I* think it will rain”, “*I* have toothache”. One can point to the difference between these two categories by saying: the cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it, the possibility of an error has been provided for. …On the other hand, there is no question of recognizing a person when I have a toothache. To ask, ‘are you sure it is you who have pains?’ would be nonsensical. (pp. 66-7)

In other words, I can make reference to myself absent knowledge of particular features of myself. The two ways of using the term “I” translate into the two ways that I am conscious of myself. Here I diverge from Wittgenstein, who maintained that the use as subject was a

grammatical peculiarity of the first-person pronoun, and did not refer to anything real. I maintain that use as subject translates into an important difference in the ways that I can be conscious of myself. Namely, I can be conscious of myself not only as a set of particular objects of experience, but as the subject of those experiences.

Due to the peculiar nature of reference in those cases Wittgenstein identified as "use as subject," the inability to make mistakes about who I am referring to in those cases has led philosophers to highlight this feature as the defining characteristic of SAS consciousness, a feature that Shoemaker has termed "immunity to error through misidentification relative to the first-person pronoun," (1968). This lengthy phrase is often shortened to 'IEM' in the literature, and I will adopt this convention henceforth.

It is worthwhile to explore IEM, as it is frequently mentioned in the literature, and may provide some insight into how the examination of self-reference can go beyond the conceptual realm. To make an error of misidentification with respect to the first-person is to make a judgment about myself, take that judgment to be about myself, and be wrong. When I refer to myself as subject (SAS), I am *immune* to such an error, for reasons that I will explain below (Shoemaker, 1968; Evans, 1982). IEM does not hold for SAO, as the following example demonstrates. Daprati *et al*.'s (1997) study on self-attribution (see *Empirical Support*, below for more detail) involves hiding the participant's arm from view, and then having her perform an action with the hidden arm (wearing a glove). The participant is provided with indirect means of viewing her arm, such as a video display. In some manipulations, the arm that the participant sees is not her own arm, but the arm of the experimenter, also gloved. If the participant were to look at the display during such a manipulation, and judge that it is her own arm she is seeing, she

would be making an error of misidentification with respect to the first-person. What she takes - by visual inspection alone - to be part of her SAO, is in fact part of someone else.

In contrast, if I say, "I feel hungry," I cannot be wrong about *who* is feeling hungry, that is, the subject (SAS) to whom "I" refers, even if I am wrong about being hungry (I may have some kind of stomach upset that feels similar to hunger pangs). Shoemaker (1968) argues that there are a whole class of mental states (like being hungry), that he labels P* predicates, that we cannot err in identifying as ours. Evans (1982) expands this to include proprioception, the sense of where our body parts are located with respect to us. My position is that any predicate about an *experience* is IEM, whereas any predicate about a non-phenomenal state of affairs will fail to be IEM *in principle*. I can be wrong about whose arm I see in a display, just in case the arm in the display visually matches what I've grown accustomed to call my arm. I am wrong about SAO, but I do not make an error with respect to SAS because I cannot be wrong about my having the experience of seeing an arm in the display. I may even be in error about the modality of the experience that I am describing without violating IEM, if for example I say "I hear a thumping from the floor below," but I am really just feeling a vibration through my feet. The take-home message is that *I cannot be wrong about who is having the current experience*.

While I agree that SAS consciousness makes such immunity possible, I do not agree that immunity to error is *the* defining characteristic of SAS consciousness. Rather, immunity to error through misidentification is symptomatic of the really special feature of SAS consciousness, namely that when I refer to SAS I do so without attributing any particular feature to myself. Hereafter, I refer to this property of SAS consciousness as *non-attributive reference to SAS*. I explain this idea in more detail in the sections to follow.

The two forms of self-consciousness, SAS and SAO consciousness, are closely intertwined and difficult to tease apart. The ability to attribute any features whatsoever *to oneself* - that is, being conscious of SAO - presupposes SAS consciousness. To see why this is so, consider John Perry's "Lost Lingens" thought experiment. Rudolf Lingens is an amnesiac with no memory of his former self (including his name) who finds himself lost in the Stanford library. One of the books he happens upon is a detailed biography of a man named Rudolf Lingens. No matter how detailed the description is, however, there is nothing in the description itself that can tell Lingens that it is a description of *him*. This requires that he takes the further step of coming to the realization, "*I* am Rudolf Lingens." (1977, p. 492)

SAS, then, is the foundation upon which SAO rests. SAS is that to which I attribute the features comprising SAO, and is necessary for me to make the conceptual distinction between self and other. To form a concept of myself as an object, I need some way of categorizing features into at least two groups: (a) those that I identify as part of me, and (b) those that I identify as part of the world around me. When I identify the arm in my field of view as *my* arm, I am attributing the arm to myself (i.e., SAS), rather than to someone else. To attribute anything *to* myself, I first need a sense of myself as the subject of experience, connected in a special way to particular features (SAO). Note that even simple organisms have a means of distinguishing their own bodies from the external world, but this need not manifest itself as consciousness of self *as* subject.

As suggested in the previous paragraph, SAS consciousness is coterminous with the sense of self as *owner* of particular experiences, and at this stage it is important to raise the fine distinction between ownership and authorship, because the sense of self as the *doer*, that is, the agent or author of one's s thoughts is often mistakenly taken to be identical to the sense of self as

knower or owner. While we do not ordinarily notice the difference, there are instances where we can be mistaken about the origin of our actions, as shown in Daprati *et al.* (1997) and Jeannerod and Pacherie (2004). We are never mistaken about the ownership of our present experience, however. In the case of mistakes with respect to agency, it becomes important to recognize a distinction between ownership and authorship/agency, lest we mistake misattribution of actions for a violation of IEM. I return to this topic in the discussion on thought insertion, below, as the phenomenon of thought insertion provides a clear example of the contrast.

SAS consciousness alone contains little content for introspection. Once I have stripped away bodily and mental states, what remains is the bare notion of the "I," that is, the *common subject* of experience (Kant, 1781/1998, A350). Indeed, one of the defining features of SAS is its qualitative emptiness, and it is what makes study of SAS, empirical or otherwise, so difficult. By contrast, I am conscious of SAO only via particular experiences. SAO consciousness is only possible when there is some self-descriptive content available, whereas I can become conscious of SAS even when such content cannot be called upon.

Owing to its emptiness of quality, some have suggested that the self is a fiction. David Hume (1739/1978) famously noted "...when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other... I can never catch *myself* at any time without a perception, and never can observe any thing but the perception..." (p. 252). This suggests that there is no such thing as the self, independent of particular experiences. The present study, however, is concerned with *consciousness of* self, and makes no commitments to the existence of the self as a particular kind of entity. I can sensibly distinguish the experience of being an SAS from other forms of experience (like perceiving features of SAO) without committing to any ontological claims about what I am in fact conscious of, just as I can study

perception without committing to any claims about the reality of the objects that I perceive (I could be viewing an optical illusion, or experiencing an hallucination, for example). There may be no persistent entity at the core of my experience; what is significant about self-consciousness is that I have SAS consciousness, even if the *me* of ten years (or even ten minutes) ago and the *me* of the present instant are in fact different entities.

Properties of SAS Consciousness

We can identify at least three distinct properties of SAS consciousness. First, reference to SAS is indexical, meaning that when we use a first-person reference, we use it as a pointer to SAS. Like other indexicals, such as "here" and "now," the referent depends on the context in which it is uttered. In the case of "I" and cognates, the referent is the person who utters the containing sentence. Importantly, a key piece of information is lost from a sentence when we substitute a (non-indexical) description or proper name for the first-person pronoun; namely, the sentence no longer captures that the speaker is also the subject of the sentence. Second, SAS is universally available in experience, meaning that it can readily be referred to regardless of the particular contents of one's experience. That is, we can become aware of SAS while having any experience whatsoever. Third, reference to SAS is non-attributive, meaning that when I refer to myself I need not attribute any particular feature or set of features to myself. This feature is of central importance, because it is what sets SAS consciousness apart from other mental phenomena. In the following sections, I give an overview of these features—and where appropriate, some discussion of the background conceptual work. In Chapter II, I will discuss these features in even more detail with respect to empirical work.

Reference to Self-as-Subject is Indexical

Brook (2001) identifies two unique features of SAS consciousness. First, we refer to SAS via first-person indexical. An indexical is a construct (often linguistic) that "points to" something, in this case, SAS. "I", "me", "my", and "mine" function as first-person indexicals, referring to the speaker. The first-person indexical, however, need not be linguistic. I can express the same idea with the physical gesture of pointing to myself.

The first-person indexical is what Perry (1979) calls an "essential" indexical, owing to its irreplaceable status (pp. 143-4). By substituting a non-indexical term for the first-person indexical in a sentence such as "I went to the store," such as "Ted went to the store," the sentence no longer captures the sense that the grammatical subject is the speaker of the sentence.

Perry recounts an incident where he was in the supermarket following a trail of sugar, so that he could inform the owner of the torn bag. As he followed the trail up and down the aisles, he noticed that the trail was getting thicker, and realized that it was in fact him who had the torn bag. Perry believed that the shopper with the torn bag was spilling sugar on the floor, and was correct. The key piece of information that was missing was that *he* was the shopper. Once he had this information, he was able to rectify the situation (p. 146).

Recall Perry's other example, "Lost Lingens" (1977, pp. 492-3). In order for Lingens to recognize that the description is in fact a description of *him*, he would have to recognize that certain features of the description match certain features of his SAO, and infer that the similarities are great enough that the person he is reading about is him. What is important to recognize is that there is nothing in the description itself that can possibly entail that the description is of *him* (i.e., the subject experiencing said description).

These examples show that reference to SAS must be indexical, because no description or proper name can, by itself, capture the idea that the speaker (or reader, in the latter case) of the description is also the subject of the description.

Self-as-Subject is Universally Available

Secondly, SAS is universally available during any experience. By "universally available," I mean that consciousness of SAS can be triggered during *any* experience. SAS always appears as the thing having experiences, regardless of what features I happen to be conscious of at any given moment (Brook, 2001, p. 25). Were SAS tied to a particular experience or set of experiences—for example, the experience of my body—that would have to be a part of every experience in which I was conscious of myself. Note, however, that this is not to say that SAS consciousness *must* always accompany consciousness of the world or features of self.

Arguably, no other kind of experience is universally available. Compare the perception of a tree. This experience is only available to me when a tree is within my visual field, and I happen to attend to it. If I close my eyes or turn away, the perception of the tree is no longer available to me. Likewise, many internal mental states are not universally available. I may suppose that I am able to imagine myself angry, but the actual feeling of anger is only available to me when I am in certain situations, or force myself to think anger-inducing thoughts.

As Brook (2001) notes, "we can be aware of ourselves as subject just by doing acts of representing." (p. 18) Representations may have any content whatsoever; the crucial point is that SAS consciousness can occur regardless of the particular representational content at any given moment. As I mentioned when I first introduced the SAS/SAO distinction, SAS consciousness has this unique property because it is fundamentally part of the *structure* of representation, not

the contents. If SAS consciousness relied on specific contents, the act of representing alone would not be sufficient for self-awareness.

SAS consciousness enables the person to recognize herself as the common subject of all of her experiences, regardless of how different the content of those experiences may be. To illustrate this point, you need only compare some examples of very different experiences you have had, such as being on a quiet tropical beach at noon on a sunny day versus being at a loud outdoor rock concert at night during freezing weather. Even if you could find *some* common elements between these experiences, you could find some other experience that did not share those elements, with the exception of one special element–that it was *your* experience. You cannot rely upon specific sensory modalities, such as visual or auditory elements, since it is possible to have a purely auditory experience and a purely visual experience and still regard oneself as the single common subject of those experiences.

Since SAS is part of the structure of representation, it is not "experience-dividing," to borrow Bennett's (1974) terminology (see also Brook, 1994, p. 86). As Bennett explains, to say that something is not experience-dividing is to say that we can draw "no direct implications of the form 'I shall experience C rather than D'" with respect to it (1974, p. 80). What this means is that since the subject of one experience is the same subject in every other experience, you cannot compare or contrast the subject of one experience with that of another (Brook, forthcoming, §9:5).

Many if not most experiences can be divided—at least in principle. Consider the experience of standing on a beach, enjoying a sunset, for example. My experience includes a number of features, such as the appearance of the sun glinting off the rippling waves, the clouds in the distance, the sand beneath my feet. I can compare and contrast this experience with, for

example, the experience of sitting in my office on a rainy day, which has a very different set of features. More to the point, the kinds of experiences that I have of SAO are similarly divisible. I can compare or contrast the image I have of myself in my late twenties with the quite different image I had of myself at the age of ten. In contrast, the conception I have of SAS is always the same, whether at age 10 or 100.

The realization that there is a single common subject of experience just is what is meant by SAS consciousness. This is not to say that I can ever be conscious of *only* SAS, as it is difficult, as Hume noted, to think of myself without invoking some particular feature. When I close my eyes and attempt to think of myself merely as a common subject of experience, some particular representation such as the mental image of my face invariably enters my mind.

I have already mentioned that universal availability to experience does not mean that SAS consciousness is actually present during all experience. There may be situations where a person can be conscious of things in the world—including features of SAO (but not recognized as such)—yet fail to be conscious of SAS. This appears to be the case in early childhood, as I show in the *Empirical Support* section below. From birth, infants manifest many indications that they are aware of the world around them, and can readily discriminate external sources of stimulation from internal ones. However, it is not until around the age of 18 months when infants begin to demonstrate signs of self-recognition (Lewis and Ramsay, 2004). The "mark test" (where the infant is tested on whether or not she responds to a mark on herself when presented with a mirror) and correct usage of the first- and second-person pronoun are the indicators most often cited.

A person may also fail to be conscious of SAS when she is entirely focused on a task, such as when performing a complicated activity, like simultaneously keeping track of a number

of basketball players during a game (as in the famous "Gorilla" video familiar to many first-year psychology undergraduates). During such tasks, she may become so absorbed as to "forget herself," a phrase that seems particularly apt to describe the behaviour. Further study is needed to determine how literally we interpret the expression "forgetting yourself." Such behaviour offers a promising direction for the empirical study of SAS consciousness. If a person can be caught in a moment of absence of self-consciousness, then I should be able to create such an absence in an experimental setting to identify further characteristics of SAS consciousness. An experimenter could engage a participant in a task requiring a large investment of attention, for instance, then interrupt the process with a picture of the participant, which is certain to elicit self-directed thoughts (as would looking into a mirror). We can learn more about the psychological structure of self-consciousness by observing how other cognitive faculties—such as episodic memory—are affected in the absence of SAS consciousness.

Reference to Self-as-Subject is Non-Attributive

Finally, since SAS is available in any experience whatsoever, reference to it is also *non-attributive.*[1] In other words, we can refer to SAS without attributing anything in particular to SAS. This property is closely connected with the universally available nature of SAS, but deserves mention as a separate feature. Since SAS is universally available, if reference to it required attribution of a particular representation, that representation would also need to be universally available. Since there is no representation universal to experience—other than the representation of SAS—reference to SAS must be non-attributive.

It should be noted that although universal availability logically entails non-attributive reference, the reverse relation does not hold. This follows from a more general claim about

[1] What Brook (2001) calls "non-ascriptive" self-reference. I have opted to use the term "non-attributive" instead to facilitate cross-disciplinary understanding.

reference in general, namely that reference to an object is not sufficient for consciousness of said object (Brook & Raymont, forthcoming). In addition to reference, something else is needed; one suggestion is that attention must be paid to the object. For example, I may utter the phrase "*I* am away from the phone right now" while recording a message on my voicemail without necessarily attending to myself—I am familiar enough with the requirements of recording a voicemail message that I may do so while my attention is fully occupied by another task.

In practice, reference is not so easy to separate from consciousness, so non-attributive reference is often coupled with consciousness of SAS. I can be conscious of myself as a subject without being conscious of any particular properties of my bodily or mental states (Brook, 2001, p. 9), such as hair colour or features of my personality. In other words, I need not *attribute* features to myself in order to be conscious of myself. We can find mention of this feature of self-consciousness in Kant (1781/1998), who wrote that "...through the I, as a simple representation, nothing manifold is given." (B135) This nicely captures the essential point that SAS (the "I"), by itself, does not provide us with any additional content. It is a "simple" representation, meaning that it is irreducible to component parts.

If I think or say, for example, "I am seeing the colour red," I need not have in mind *any* particular properties of myself. All I can say about myself is that I am seeing the colour red, but this does not provide me with any additional information about myself. We could go further and, taking a cue from Descartes, merely assert that "I am," or "I exist." Again, nothing "manifold" is given about self when I assert such a statement.

To better understand what it means to say that reference to SAS is non-attributive, it is important to understand what reference in general is usually like. Most reference is *attributive*, meaning that when I refer to something I assign certain features to it. When I use the word "cat"

to refer to my cat, for example, I have in mind a particular feature or set of features associated with my cat. Similarly, when I use a proper name like “Ted,” I have in mind a particular person, whether that person happens to be me or someone else. Using a proper name also requires that I disambiguate the reference – I need to ask, “Who is Ted?” By contrast, when I use the word “I” as a subject, I need not have in mind any features that I associate with myself as an object. The “I” in the expression “I feel pain”, for example, is the same “I” as in “I live in Ottawa,” but there is nothing in common between the two expressions other than that they refer to the same subject of experience. When I say, “I see a tree,” I am not attributing anything to the “I,” other than that it is the subject of the experience of seeing a tree. When I describe my seeing a tree, I don't need to disambiguate reference to whom “I” refers. I don’t need to ask, for example, ‘Is it I that sees the tree, or Kylie?’ Whereas, if someone were to say “she sees a tree,” and there were several women in the vicinity of the speaker, I would need to ask ‘to whom does “she” refer?’

Now it may seem contradictory that, having just said that one can refer to SAS without attributing features to it, I proceed to mention at least one apparent feature, that it is the subject of experience. This would be a problem if a “subject of experience” were like the visual sensation of a tree, or of my image in the mirror. A “subject of experience”, however, is unique in that it is always available in experience (see the previous section), whereas the tree or self-image is not. The experience of anything whatsoever makes SAS available, whereas only certain experiences make available, for example, the image or concept of a tree or one’s body.

We have already encountered the non-attributive property of SAS consciousness above, in my discussion of the distinction between SAS and SAO, and the phenomenon of IEM. We can see how IEM follows from the claim that self-reference is non-attributive. It is in virtue of the fact that SAS is featureless, and thus cannot be contrasted with objects of experience or other

instances of SAS, that I can have IEM with respect to SAS. When I am aware of SAS, it is always presented in the same way. As such, SAS cannot be contrasted with an alternate SAS in the same moment—that is, synchronically—nor can SAS be contrasted with previous or future instances of SAS—that is, diachronically.

Once again, contrast the experience of a tree. I can distinguish one tree from another—both synchronically and diachronically—by various features, such as the shape of the leaves or the hardness of the bark. In the synchronic case, if I see two trees next to one another, I can distinguish them by their shape, number of branches, leaf density, and so forth. Even if, improbably, the trees were identical in appearance, I could still contrast them by spatial properties—the one on the left versus the one on the right, for example. In the diachronic case, if I am looking at a single tree at the present moment, and think on my experience of a tree I saw yesterday on the other side of town, I can compare the present experience of the tree with my previous experience of a tree. Even if I am confused about where I saw the tree yesterday, I may recall that the tree I saw previously was an elm, whereas the one I am currently looking at is a cedar.

There are no analogous features by which we can distinguish one subject of experience from another. I am only ever aware of one subject (namely, me) – I am only aware of others as objects, and if I were to become aware of another subject in the same way that I am aware of myself, there would be nothing to distinguish that subject from myself. I take the synchronic case to be unproblematic, but the diachronic case may be difficult to appreciate. It is the case that when I recall my experiences from yesterday, or ten years ago, I recall a number of differences between the person who had previous experiences and the person having the current experience. For example, I remember that yesterday I was wearing black dress pants, whereas today I am

wearing beige khakis. However, I have no trouble recognizing that it was *I* who was wearing the black dress pants and saw a tree on the far side of town. Indeed, because reference to the self-as-subject is non-attributive, if I were to somehow recall someone else's experiences from yesterday, there would be nothing to the *structure* of that experience that would indicate that it was someone else's experience. The only way to tell would be to contrast the *contents* of experience—for example, recognizing that it is highly improbable that I was wearing a green sweater and climbing the mountains of Tibet yesterday, knowing that I do not own a green sweater and am currently in Canada.

There is also a clear connection between the indexicality of self-reference - the first property of SAS consciousness[2] that I mentioned - and non-attributive reference to self. As Brook notes, "indexical reference to self could not be essential [i.e., irreplaceable] unless there is a way of doing such acts of reference that is independent of (non-indexical) identification" (p. 14). Shoemaker (1968) explains that identifying a particular object with self could not serve as the basis of first-person indexical statements, because I would still need to establish a relation between the thing identified and the self. The identification of any particular thing with self presupposes the featureless SAS, because I must either identify something true of the object that is also true of myself, or identify a unique relationship between the object and myself, namely that it is a part of myself. Attempting to reduce self-consciousness to identification with a feature or object would lead to an infinite regress, as every feature identified with the self would have to relate to some further feature, *ad infinitum*. At some point, I must refer to the featureless SAS in order to break out of the regress (p. 87). In identifying my body as myself, for example, I am positing that this body, and not someone else's, belongs to me. What is this "me" that my body

[2] Notice that I wrote "property of SAS *consciousness*," not "property of SAS." SAS is the "subject component" of my experience, while SAS consciousness describes the process behind that experience.

belongs to? If I further identify myself with a further particular representation, such as a mental state, I am still positing a relation between my mental state and something further, namely myself. To cease the infinite regress, I must accept the idea of "myself" *simply* as the subject of experience, full stop.

One might object that the problem of infinite regress is a red herring, using the argument that merely having a representation of me should be enough, by itself, to identify it *as* myself. It appears odd to suggest that I need, in addition to a representation, the recognition that I am the subject of that representation. I never seem to make mistakes about *who* is having the experience, therefore—my objector might argue—I do not need to posit an "I" in addition to my representations. Note, however, that the oddity of suggesting that one could be conscious of a representation not one's own stems from taking SAS consciousness for granted.

We take for granted that all first-hand experiences are our own rather than someone else's. It seems odd to claim that the experience of skydiving, for example, is someone else's experience. While it may be impossible or at least highly unlikely to have someone else's present experiences,[3] there can be deficiencies in a person's ability to attribute particular experiences to self. The case of past experience is not so clear. People can make "source confusion" errors, a particular kind of source-monitoring error, recalling in vivid detail events that they think happened to them, but that they did not actually witness. These events may be unknowingly fabricated or real, as for example when the person hears a story and later mistakenly thinks that she witnessed it personally (Schacter, Chiao, & Mitchell, 2003).

People can also make errors with respect to the source of an experience, confusing self-generated thoughts for thoughts generated by other agents, a phenomenon known as thought

[3] I will set aside a number of interesting discussions concerning this point, as they are tangential to our current concerns.

insertion. Stephens and Graham (2000) document such cases, where a person experiences certain thoughts or internal monologues as being placed in her mind by another agent. The authors describe this as a breakdown of self-consciousness, stating that when self-consciousness "breaks down or becomes disturbed, it appears to the self-conscious person as if *other* selves or agents are involved in his or her stream of consciousness." (p. 2) People who experience thought insertion often suffer from schizophrenia, though the phenomenon is not restricted to abnormal psychology cases. Nor is misattribution of mental phenomena restricted to thought insertion, as some people experience 'alien voices,' that is, verbal hallucinations where the person attributes the voices to another person.

It is important to note that such phenomena do not violate IEM, because IEM as I have cast it only claims immunity for experiences that I claim as *mine*, not someone else's. Specifically, I cannot have an experience, claim that it is mine, and be wrong. In the cases just mentioned, patients claim that the experience of another has been inserted into their minds, when those thoughts were actually generated by them. They have made an error with respect to the *third*-person. Furthermore, the error is not a misattribution of ownership, but of agency. Even in the case of inserted thoughts, the patient recognizes that *she* is having the thought; she is just mistaken about its origin. Stephens and Graham (2000, pp. 154-5) claim that we can make sense of this if we recognize that the self as *subject*—or owner—of the experience (a thought, in this case) can be dissociated, albeit in abnormal conditions, from the self as *agent* or author – the originator of the experience. This distinction is parallel to the SAS/SAO distinction I am making here, although sense of SAO encompasses far more than the sense of agency. The cases of inserted thoughts, as well as the cases of misattribution of action reported by Jeannerod and Pacherie, suggest that agency is an aspect of SAO. Agency is part of the *content* of experience,

because it is something that can be attributed *to* self, and can be contrasted with other sources of agency.

Summary

I have identified three key features of the mechanism behind SAS consciousness. First, reference to SAS is indexical. Second, SAS is available during any experience whatsoever. Third, reference to SAS is non-attributive. The next step is to determine how these features make a difference to the behaviour of an organism. One way to do this is to identify tasks that require some or all of these features. Consistently correct use of the first-person pronoun, for example, is possible only when one appreciates that the pronoun is indexical to the speaker. The next section is a review of the experimental work in cognitive psychology and neuroscience. Following this, I describe a pair of my own studies designed to examine the capacity for SAS consciousness.

Empirical Support for SAS/SAO Distinction

Introduction

In this section, I provide a critical review of existing studies of self-consciousness, as they pertain to establishing empirical evidence for the distinction between SAS and SAO. In addition, I highlight how those researchers that fail to acknowledge the distinction generate implausible interpretations of their results, and I will suggest alternative interpretations made possible by acknowledging the distinction.

The purpose of this section is to establish a crucial part of the groundwork for my own studies, as there I will take the SAS/SAO distinction as a basic assumption. Here, I will demonstrate that an abundance of evidence for the SAS/SAO distinction is available, such that I do not need to do further studies to support my use of said distinction.

In the *Research on Indicators of SAS Consciousness* section in the second half of this work, I will review what little is understood about the cognitive underpinnings of SAS consciousness. While my own empirical work will be restricted to the methods of cognitive psychology, I will also review some of the relevant work in cognitive neuroscience. In the *Empirical Studies* section of Chapter II, I will give a detailed description of my own studies, with the aim of addressing the second goal of my project, namely furthering knowledge of the psychological mechanism of SAS consciousness. I study the effects of inattention on SAS consciousness, testing my claim that we can be conscious without being conscious of SAS. I also investigate the development of SAS consciousness in young children (three- to five-year-olds), specifically how they learn to integrate remembered past instances of self with the current sense of self. Due to sample size constraints, the developmental study was conducted as a preliminary investigation.

Given the plethora of philosophers who recognize a fundamental distinction between subject and object, it is perhaps surprising that psychologists who study the self rarely take notice of the distinction. As mentioned earlier, much of the psychological literature focuses on the self-concept, and either fails to recognize SAS or takes it for granted. It is thus necessary to show that the distinction is not merely philosophical speculation but has some empirical basis.

Before I begin the review, I must address a terminological issue. SAS is sometimes referred to as the "pre-reflective" self, in contrast with the self as an object of reflection. This is the terminology that Legrand (2007) uses, as does Rochat (2003) and Rochat and Striano (2000). I have chosen not to use this terminology, however, because it is not clear whether or not "pre-reflective" is an accurate description of the phenomenon in question. Furthermore, in the case of

Rochat and colleagues, it is not clear that what is meant by "pre-reflective" self is parallel to my definition of SAS.

There are a small number of empirical researchers who recognize the distinction between SAS and SAO (of particular interest are Jeannerod & Pacherie, 2004; Legrand, 2007; Legrand & Ruby, 2009; but see also Travis, Arenander & Dubois, 2004; Travis, 2006; Tsakiris, 2010). Notably, William James (1890) distinguished the "self as subject," or self as knower, from the self as object or self as known. The distinction has fallen out of use in empirical research until very recently, however. Fast, Marsden, Cohen, Heard and Kruse (1996) highlight the fact that most quantitative studies of self focus on SAO. As they note, "[m]otivational forces are closely associated with self-images or goals but the dynamic processes themselves, *praising* or *criticizing* oneself, *working* to achieve a goal, *avoiding* the actualization of an unwanted possible self, or *activating* a particular self-image are not identified as self aspects" (Fast *et al*., 1996, author's emphasis, p. 34). The "dynamic processes" they identify as the self as subject, citing the theoretical use of the term in James, psychoanalytic theory, and Piaget.

Many researchers do not explicitly mention the distinction, but appear to take it for granted. That is, without assuming the distinction between SAS and SAO, the claims made by these researchers would make little sense. The category includes almost all research on errors of self-attribution (i.e., misattribution of bodily or mental states *to others*),[4] most notably Daprati *et al.* (1997), Petkova and Ehrsson (2008) and Lenggenhager, Mouthon and Blanke (2009). I will examine these cases in more detail below.

Other researchers fail to recognize the distinction between SAS and SAO—even implicitly—introducing methodological difficulties and faulty interpretations of results. I argue

[4] Not to be confused with the kind of misattribution we are immune to (see discussion of IEM), which are misattributions *to self*.

that by keeping the distinction in mind, we can make better sense of the results of these studies. In particular, I will examine the work of Rochat (2003; Rochat & Striano, 2000) on the purported development of self-awareness in infancy, and Povinelli's (2001) review of his work on self-recognition in children. I will take a more detailed look at alternate interpretations of Povinelli's work, as it is particularly relevant to my preliminary developmental study.

Empirical Evidence for the SAS/SAO Distinction

Few researchers explicitly mention the distinction between SAS and SAO, either taking it for granted, failing to recognize it as a relevant distinction, or using alternative distinctions, such as Rochat's five levels of self-awareness (2003). Here I present a few exceptions, namely the work of Jeannerod and Pacherie (2004), Legrand (2007; Legrand & Ruby, 2009), and Tsakiris (2010). Legrand *et al.* do not present their own studies, but provide alternative interpretations to others' studies in light of the SAS/SAO distinction. Jeannerod and Pacherie (2004) make use of the distinction in claiming that misattribution errors can be made with respect to SAS, thereby violating IEM. While I disagree with this claim, Jeannerod's work is an important instance of the relevant distinction at work. Tsakiris (2010) distinguishes between a sense of agency and ownership, focusing exclusively on the sense of body-ownership. In particular, he looks at studies on the "rubber hand illusion," where people develop a sense of ownership for a false limb under certain conditions. I will revisit this illusion and similar experimental manipulations in the course of this review.

In support of this distinction, I review the literature on dissociations between sense of ownership and sense of authorship or agency (Synofzik, Vosgerau, & Newen, 2008), including a condition known as asomatognosia (Heydrich, Dieguez, Grunwald, Seeck & Blanke, 2010), where the patient lacks a sense of ownership over significant portions of her body while still

maintaining motor control over the affected areas. I will also review a study on the everyday phenomenon of "absent-minded action-slips" (Cheyne, Carriere & Smilek, 2009), where people perform the opposite of an intended action—such as taking a candy out of a wrapper and throwing the candy in the trash instead of the wrapper. The authors found that when a person is made aware of an action/intention mismatch while the action is being performed, the person experiences a sense of alienation of agency.

I will also look at studies that support the distinction, but do not explicitly mention it. Much of the research focuses on the attribution and misattribution of actions. Daprati *et al.* (1997) examine the differences between normal and schizophrenic patients on the ability to correctly attribute a hand presented on a television screen to self. Petkova and Ehrsson (2008) extend the research on self-attribution from single limbs to the entire body. Similarly, Lenggenhager, Mouthon and Blanke (2009) did a nearly identical study, using the same apparatus but a different measure. Together, these studies demonstrate that people can be wrong at least with respect to all bodily components of SAO, yet retain a clear sense of SAS.

The aforementioned studies are specific to bodily attributions to self. The SAS/SAO distinction is also implicitly recognized in studies on autobiographical memory, particularly episodic autobiographical memory, and episodic memory in general. I will provide an overview of the relevant literature on episodic memory later. For the purposes of demonstrating the implicit use of SAS/SAO, I will look at Addis and Tippet (2004) and Piolino, Desgranges and Eustache (2009). The cases that they report—Alzheimer's and other disruptions of episodic autobiographical memory (EAM)—support the distinction by demonstrating that people who are incapable of remembering anything about themselves nevertheless retain a sense of self, albeit

confined to the present. This has also been demonstrated in the well-known case of H.M., and the lesser-known case of K.C.,[5] as explained in Tulving (2002).

Dissociations Between Ownership and Authorship

Observed dissociations between sense of ownership and sense of authorship (i.e., agency) in patients afflicted with "alien voices" and inserted thoughts provide further reason to make use of the SAS/SAO distinction. Recall the claim by Stephens and Graham (2000) that one's sense of self as an agent is closely connected to, but not equivalent to, sense of self as the subject of experience, that is, the owner of the experience. A patient can be mistaken about the origin of her internal 'actions' (e.g., thoughts and inner speech), yet recognize that *she* is having the experience—the thought occurs *in her mind* although it is initiated elsewhere. Such cases show dissociation between agency and ownership,[6] and it is precisely such dissociation that lends to the disturbing quality of the experience. I regularly attribute thoughts to myself and to others—I imagine that it would be unsettling if what appeared to be the thoughts of others, unmediated by (external) voice, were to enter my awareness.

On my account, since agency is the attribution of action *to* self, we should expect it to be a component of SAO and therefore separable from SAS at least in principle. That there appear to be instances of such separation in practice lends further support to the idea. In addition, attribution of agency, as with any other attributive reference to self, can be subject to errors of misidentification. One can experience an action, think it to be their own, and yet be wrong. Many of the misattribution studies that I will consider in the next section (*Implicit Use of the SAS/SAO*

---

[5] I will return to the case of K.C. several times throughout this work.

[6] Importantly, however, such mistakes are *not* counter to IEM with respect to the first-person because the subject is attributing her own action to someone else, and IEM is required only to hold in the opposite direction. As we will see, actions *are* subject to errors of misidentification. That is not the type of error being made in the cases of thought insertion and alien voices, however.

Distinction) reveal that misattribution of agency is not difficult to elicit under experimental conditions. Anecdotal evidence suggests that such errors frequently occur in everyday situations when events are strongly correlated with one's intentions.

Synofzik *et al.* (2008) propose a framework for studies of the "acting self," which they divide into two parts, sense of ownership (SoO) and sense of agency (SoA). The SoO/SoA distinction is parallel to the SAS/SAO distinction that Legrand and I recognize, but is narrower in scope, focusing on action alone. Based on their framework, the authors contrast two similar disorders that are often conflated, alien limb syndrome (sometimes called alien hand syndrome or AHS, because it frequently affects the hand alone) and anarchic limb syndrome,[7] as disturbances in sense of ownership and sense of agency, respectively (p. 421).

Briefly, alien limb (alternatively, alien hand) syndrome is a condition whereby the patient feels as if she is not the *owner* of a limb. Cases of alien limb syndrome are not normally associated with involuntary movements of the limb, so the person does not experience alienation of agency. In many instances, the person has propositional knowledge of ownership, but competing information at the level of feeling (as opposed to the level of judgment). In some cases, however, the person can develop delusions concerning the owner of the limb, such as believing that the limb belongs to a parent. In one instance, a woman expressed surprise upon seeing that the "alien" hand was wearing her rings (Synofzik *et al.*, 2008, p. 421).

Anarchic limb syndrome differs from alien limb syndrome in that the patient does not feel, or develop delusions, that the limb *belongs* to someone else, but might refer to the limb as an autonomous agent. Again, one can have propositional knowledge of the agency, but fail to

---

[7] In much of the literature, no distinction is made between *alien* and *anarchic* hand syndrome, in large part because the distinction is based entirely on the agency/ownership distinction, which is often not recognized (Synofzik *et al.*, 2008).

register that agency as a feeling. As with alien limb syndrome, it is possible to develop delusions about who is controlling the limb (Synofzik *et al.*, 2008, p. 422), paralleling the cases of thought insertion and alien voices documented in Stephens and Graham (2000).

Alienation of agency can also occur with respect to nearly the entire body without corresponding lack of control. Heydrich *et al*. (2010) claim that the congruence of visual and other sensory inputs are essential to self-location, first-person perspective, and most importantly for present purposes, self-identification. In cases of global—that is, across the whole body—illusory "own body" perceptions, patients experience abnormalities in these three faculties. Heydrich *et al.* studied two cases of epilepsy where patients experienced abnormal self-identification, reporting numbness and feelings of alienation from a significant portion of their bodies. During seizures— which occurred daily and typically lasted about a minute—Patient 1 reported that the entire left side of his body, including the head, felt as if occupied "by a stranger" and disconnected from the right side, where he located self. These seizures were only partial, and he could perform many tasks such as walking and talking during them (Heydrich *et al.*, 2010).

Patient 2 experienced a gradual onset, with the sensation of alienation beginning in the legs and lower trunk and progressing all the way up to the neck at the height of seizure. He reported that his entire body from the neck down (arms excepted) felt numb, weak, and disconnected from the head. Interestingly, he also felt disconnected from his thoughts and his past, and felt that he did not have control over his actions and speech. He also felt disconnected from the environment, with perceptions appearing "transformed and distant" (Heydrich *et al*., 2010, p. 706). Neuropsychological tests revealed that he became highly distractible and experienced intrusive thoughts during episodes (Heydrich *et al*., 2010). This last observation is

particularly interesting because it suggests a link between attention and self-awareness, a link that I will explore later in this review.

Less extreme—but easier to study—instances of alienation of agency can be invoked in mentally-healthy people. Cheyne, Carriere and Smilek (2009) studied so-called "absent-minded action-slips" (p. 482), where the body automatically performs certain actions contrary to one's intentions. An anecdotal example is the frustrating experience of taking a candy out of its wrapper and mistakenly throwing the candy in the trash, rather than the wrapper. They describe such an occurrence as a failure of "self-awareness during instrumental action." (p. 481) Such actions can be complex, goal-directed behaviours such as going into the bathroom with the intent to take medication and finding yourself brushing your teeth (Cheyne *et al*., 2009).

They found that under certain conditions, absent-minded action-slips are often associated with a feeling of "alienation of agency," as if the body were a separate agent acting independently of self. An anarchic limb syndrome (what Cheyne *et al.* call anarchic hand syndrome) patient will regularly experience her hand acting as if it were controlled by another, even though she is not delusional and recognizes that the hand is actually her own. Absent-minded action slips are similar in kind but differ in duration. Often, we catch these intention-action mismatches after the fact, resulting in—at most—a feeling of embarrassment. In order to experience alienation of agency (the sense that the action is being controlled by another agent) the mismatch between intention and action must be caught *during* the action (Cheyne *et al*., 2009).

To induce absent-minded action-slips, Cheyne *et al.* (2009) used a variation on the SART (Sustained Attention to Response Task), which was designed to measure behaviour associated with mind-wandering. The SART is a type of inhibition task—that is, the goal is to *inhibit* a

response. In the modified SART, the participant was instructed to press a button unless an infrequent number appeared. This task was chosen after noticing in an earlier experiment that many participants expressed frustration when they intended to withhold yet pushed the button. In the 2009 study, participants frequently had difficulty withholding during trials where they were instructed to withhold, despite having no difficulty identifying the target. Cheyne *et al.* (2009) note from personal experience and the reports of participants that people often seem to recognize the error even while performing it.

Cheyne *et al.*'s work provides further evidence for dissociation between sense of SAO (agency, in this case) and sense of SAS (i.e., ownership). Furthermore, that such dissociations can occur in mentally-healthy individuals demonstrates that they are not due to abnormalities in the structure of experience. Rather, syndromes like alien and anarchic limb and the rare cases of epilepsy cited above demonstrate that the normally unnoticed SAS/SAO distinction can become readily apparent when the ordinary mechanisms linking the two break down.

Jeannerod and Pacherie (2004) explicitly mention the SAS/SAO distinction, stating that "the problem of self-identification does arise, not just for the self as object, but also for the self as subject." (p. 114) Here, the authors are making the surprising claim that SAS is *not* immune to errors through misidentification with respect to the first-person (IEM), an idea that I introduced in the *Conceptual* section. It is worthwhile to explore this claim in more detail, as it directly challenges a philosophical premise long thought to be irrefutable—that I cannot be wrong about being the subject of my experience. I will argue, however, that the authors fail to show the susceptibility of SAS to errors of misidentification with respect to the first-person.

The authors correctly note that "the problem of self-identification can … only arise for those forms of self-ascription that are not identification-free" (p. 114). SAS is supposedly

identification-free, so it should not be susceptible to such a problem. The authors' claim that SAS fails to be IEM, then, hinges on the premise that SAS is not identification-free (p. 117). This premise is, in turn, based on some controversial claims about the nature of intentions, and how we determine the author of an intention, which I expand on below. It is also based on the premise that agency—specifically authorship of intentions—is inextricably linked with the feeling of ownership (sense of SAS), such that an error with respect to the former is an error with respect to the latter (p. 123).

Jeannerod and Pacherie (2004) claim that the sense of ownership of one's own body is established via two mechanisms, a) the matching of visual, tactile, and proprioceptive signals with each other, and b) the matching of one's intentions with one's actions (p. 117). Concerning (a), Jeannerod and Pacherie recognize a subtle distinction between proprioception and perception of tactile information. Proprioception provides information about *where* the body part is in relation to the rest of the body, while tactile information tells us *what* body part is being represented (p. 118).

The authors note that visual information about ownership can easily override the tactile and proprioceptive information, such that even when proprioceptive information tells you that your arm is in one location, if the arm visually appears to be in a conflicting location—within certain constraints that I will elaborate on later—you will feel as if the arm is in the visual location. This effect was previously demonstrated in Botvinick and Cohen (1998) using the "rubber hand illusion" paradigm. In the rubber hand illusion, a participant's arm is concealed and a rubber arm is placed in front of her. In one condition, the experimenter strokes both the participant's hand and the rubber hand in synchrony. In this condition, Botvinick and Cohen (1998) found that participants felt the touch where it was seen on the rubber hand, and not where

it was actually touched, several inches away on the real hand.[8] When the strokes are applied asynchronously, the illusion does not occur.

The rubber hand illusion is of interest for my project, because it demonstrates that even though I can make errors regarding where my body parts are in relation to other parts—for example when misleading visual information overrides proprioception—I am not thereby mistaken about being the owner of the sensation, and IEM is not violated.

Claim (b)—that sense of ownership of the body is based on the matching of intentions with actions—is based on the following considerations:

1) We are aware of the intentions of ourselves and others non-inferentially, that is, we perceive them directly via the process of action simulation. The same parts of the brain that are active when we plan our own actions are active when we observe someone else's actions.
2) We are aware of intentions as "naked" intentions, that is, there is nothing in awareness of an intention to identify whose intention it is. An additional step is needed to identify the owner of the intention.
3) The process by which we attribute an intention to an owner is not entirely reliable, so it is possible to misattribute in certain circumstances.

According to the authors, action simulation is the primary mechanism by which we self-attribute. On the action simulation theory, simulated and actual actions are indistinguishable at the neural level save for the execution. A person is better at simulating her own action than simulating that of another, because she is far more familiar with her own actions. However, by

[8] It should be noted that there are a number of prerequisites needed to elicit this effect; the rubber hand must be correctly aligned anatomically, must be within reaching distance, and the brush strokes must be applied in the same direction (Costantini & Haggard, 2007).

manipulating the accuracy of simulation, the experimenter can cause a person to identify herself as the source of someone else's action. The authors claim that we use a non-inferential process– by that they mean a *direct*, unmediated process, which they equate with simulation–when we attribute intentions to an agent, whether that agent is self or other (p. 139). When the simulation of another's action is sufficiently similar to the simulation of my own actions, then it is possible to confuse them. Since I can be confused with respect to my own agency—and on the authors' account, agency is the primary form of self-identification—the authors reason that other kinds of non-inferential information, like pain and proprioception, could be prone to error as well.

The authors claim that the Alien Hand paradigm (Fourneret & Jeannerod, 1998; Nielsen, 1963) demonstrates how such errors of misidentification can be caused in mentally healthy people, and that we have a tendency to over-attribute actions to ourselves. In this experiment, the participant places her arm through a hole in a box, and is instructed to draw a line towards the top of a page. A small window in the top of the box allows the participant to observe her line-drawing. Unbeknownst to the participant, the box is fitted with a mirror so that the participant actually sees the hand of the experimenter (both the participant and the experimenter wear a glove to mask obvious differences between the hands).

When the participant is asked to draw the line, the experimenter mimics the action. Midway through the task, the experimenter makes the line diverge from straight according to a set number of degrees up to 10°. Even though the line starts to diverge, the participant still believes that she is the one drawing the line, so long as the divergence is not too great. Furthermore, she attempts to correct for the path of the line, but does not realize that she is doing so. When confronted with the discrepancy, participants commonly confabulate, claiming for instance that they must have been tired or inattentive (Jeannerod & Pacherie, 2004).

As Jeannerod and Pacherie explain, when we make a determination about who is performing an action, we rely on visual information so long as the discrepancy is not too large. So long as our intention is roughly congruent with the perceived action, we attribute the action to ourselves. The authors claim that the feeling of ownership is tightly bound to the feeling of agency, such that when we make a mistake with respect to the latter, we also make a mistake with respect to ownership (p. 123). In other words, we make an error of misidentification with respect to SAS.[9]

There are a number of problems with the authors' conclusion. First, their argument depends on the highly dubious suggestion that we are aware of intentions non-inferentially. They conflate perception of action with awareness of intention but are not justified in doing so based on the evidence they provide. There is also theoretical reason to be cautious of the conclusion that simulation is indicative of a non-inferential process, since as Dennett (1981) observes, simulation still requires the construction of a theory and is therefore inferential. The logic is as follows: when I simulate the *x* of another—where *x* could be a belief, or an action, or any sort of mental or physical state—my knowledge of the other is necessary to 'run' the simulation and that knowledge must in turn be organized into a theory (p. 79).

Legrand and Ruby (2009) give a more specific, neurological criticism of simulation theory, claiming that while simulation theory does a good job at explaining the sensory, motor, and emotional activations in an observer, the simulation theory is not adequate to explain so-called "mind-reading," (predicting another's intentions or actions) and there is no evidence to

[9] Curiously and confusingly, the authors refer to the sense of ownership as "self as object" and contrast it with "self as agent," which they take to be the self as subject. This is precisely the reverse of Stephens and Graham's (2000) distinction. Nevertheless, since they maintain that a violation of IEM with respect to the object constitutes a violation of IEM with respect to the subject, it does not matter for my purposes what side of the distinction they label as subject

suggest that the brain regions involved are particular to self-representation. The particular brain regions in question are involved in many cognitive tasks not involving representation of either self or other (p. 267).

Secondly, and more importantly for my purposes, being wrong about the author of an action is not the same as being wrong about the owner of an intention. Clearly, the participant is confused about the former, and does not have IEM with respect to actions (this should not be surprising), but actions, as movements of objects of experience, are part of SAO, not SAS. As the observations on inserted thoughts by Stephens and Graham demonstrate, one can be confused about one's agency yet retain SAS consciousness. Although cases of inserted thoughts are highly atypical, they demonstrate that such confusion is possible. In the cases that Jeannerod and Pacherie consider, the participant is not confused about her intention just because she mis-attributes the action; it just happens that someone else's action tracks that intention accurately. The problem of mis- or over-attribution reduces to a mere problem of causation, and does not introduce any difficulties for IEM.

Finally, even if the participant was to mis-attribute the action of another to her, this would not count as a counter-example to IEM. Shoemaker never claimed that IEM applied in the reverse direction—that a person could not make a judgment about someone else, think that it was a judgment about someone else, and yet think the judgment is about her. IEM only applies to judgments with respect to the self.

Though there are conceptual faults in Jeannerod and Pacherie's study with respect to their claims about IEM, their findings indicate that action recognition is the predominant form of self-identification (p. 124-5). Consider the prototypical case of action recognition, movement in a mirror. The best way to tell if an image is a reflection is to move and observe whether or not the

movement in the image corresponds to proprioceptive feedback about the positions of one's limbs. Participants in Jeannerod and Pacherie's studies had a tendency to over-attribute actions, meaning that when they were presented with visual information about the experimenter's actions concurrent with their own felt actions, they tended to attribute the experimenter's hand to self (p. 122). This tendency suggests that the sense of self as an agent (a component of SAO) is closely tied to the sense of self as owner (i.e., SAS). Over-attribution, however, is not a case of misidentification with respect to SAS, only with respect to SAO.

Implicit Use of the SAS/SAO Distinction

Evidence for the existence of the distinction between SAO and SAS consciousness can be seen in cases where people mistakenly attribute features of themselves to others, and vice versa. Such cases suggest a discontinuity between the particular features associated with myself and SAS consciousness.

Much of the empirical work on self-attribution has focused on the attribution of *actions* to oneself rather than another. Actions are properly considered features of SAO. These studies are predominantly concerned with agency, or the sense of self as a locus of control. The sense of self as *agent* is orthogonal to the current study; what is important for my purposes is that a mistake about agency is a particular form of self-attribution error, and one that can be readily invoked in an experimental setting. Such errors demonstrate dissociation between SAS and SAO. This dissociation is rarely made explicit, however, in self-attribution studies.

In a study by Daprati, Franck, Georgieff, Proust, Pacherie, Dalery, and Jeannerod (1997), for instance, the experimenters gave mentally-healthy participants indirect visual feedback (via computer display) of a simple hand movement that they had asked the participants to perform. In some trials, they replaced this feedback with the feedback from an experimenter performing an

action (either the same or different). Both the experimenter's and participant's hands were gloved to prevent the participant from relying on visual cues to determine the owner of the hand. After the trials, the experimenter asked the participant whether or not the action performed was the participant's own. When the actions performed by the experimenter were similar to the actions performed by the participant, participants made the wrong attribution in 30% of cases, while participants always made the correct attribution when the actions performed by the experimenter were clearly different.

As Legrand (2007) explains, this experiment is supposed to show that there is a distinction between being an agent and reporting that one is an agent, and further "that consciousness of one's own actions requires the ability to consciously report being the agent of the action" (p. 593). Legrand maintains that this conclusion would not be warranted if we could reduce self-recognition to judgments about SAO, because the participant cannot simply rely on any particular object of representation to distinguish one's own actions from the actions of the experimenter. The participant must also experience herself as subject to make the distinction, flawed or not (pp. 593-4). In Legrand's words, "it only makes sense to ask the subject whether she recognizes her action as her own if it is presupposed that she already experiences herself as acting… [that] implies that self-recognition presupposes pre-reflective consciousness of the self-as-subject" (p. 593). In other words, in order to *mis*-attribute features to herself, she must first have a sense of who it is that she's attributing features *to*, namely SAS. So by the nature of the question posed to participants, the researchers are making implicit use of the distinction between SAS and SAO.

One might object that a study that looks only at the self-attribution of one body part would not demonstrate the need for a distinction between SAS and SAO, on the grounds that

perhaps the underlying sense of self is the sense of one's body as a whole. Petkova and Ehrsson (2008) describes an experiment designed to test whether or not the illusion of owning a part of someone else's body—as demonstrated by Botvinick and Cohen's (1998) "rubber hand illusion"—could be extended to the entire body, an illusion dubbed the "body-swap illusion". They hypothesized that by changing a participant's visual perspective and providing appropriate sensory feedback, the participant could be fooled into taking ownership of someone (or something) else's body (p. 1). The experimenters had the participants wear a head mounted display that relayed visual information from a stereo set of video cameras. In some of the experiments, the cameras were mounted on the head of a mannequin, pointed down at the feet, while the participant was instructed to point her head towards the floor. This setup made it appear as if the participant was looking through the mannequin's "eyes." The experimenters used a rod to lightly stroke both the mannequin's and participant's torso, either at the same time (synchronous condition) or offset (asynchronous condition). In another set of trials, a knife was used to "threaten" the mannequin, to the point of being made to appear to slice into the mannequin.[10]

The measurement used in the first experiment was a simple questionnaire that the participants filled out afterward. Most of the participants in the synchronous condition reported feeling ownership over the mannequin's body, whereas none in the asynchronous condition reported such a feeling. In the remaining four experiments, skin conductance response (SCR) was measured, based on the idea a participant would perspire more if she felt the body as her own that if she merely empathized with the mannequin. Overall, the study showed that the SCRs

[10] In this experiment, the knife was drawn through a gap invisible from the participant's vantage point (Petkova & Ehrsson, 2008).

of participants were significantly higher in the synchronous condition than in the asynchronous condition, and when the "false" body was sufficiently similar to a human body. Both the questionnaire and SCR data indicated that a participant could be fooled into *feeling* as if the false body was her own—even though she *knew* it was not—provided that the visual discrepancies were not too jarring (Petkova & Ehrsson, 2008, pp. 5-6).

Lenggenhager, Mouthon and Blanke (2009) describes a study nearly identical to Petkova and Ehrsson's (2008), except that the authors quantified over spatial location instead of SCRs. Their results indicate that people locate self (i.e., SAO, though they do not use this terminology) where the touch is seen, rather than where it is felt, again demonstrating that visual information can override tactile information about self-attribution.

Aside from their recognition of the SAS/SAO distinction, the significance of the above studies is that visual perception plays a key role in the sense of ownership of my body, but only when combined with appropriate proprioceptive feedback. This suggests that a crucial mechanism for the self-attribution of features is the coincidence of proprioceptive and perceptual feedback, because when the coincidence is absent, I do not make a self-attribution.

Studies that Fail to Recognize the SAS/SAO Distinction

The studies in the previous section made implicit use of the SAS/SAO distinction. Since the focus of the studies was on the attribution of features *to* SAS, they were squarely concerned with how one goes about constructing a sense of SAO, and the results of those studies give little insight into the mechanism behind SAS consciousness.

In this section, I will examine two studies where failure to recognize the distinction between SAS and SAO affects the interpretation of the results. I will argue that recognizing the distinction in question allows for interpretations that better fit the results, and allows me to

develop further research questions that I can use to investigate the nature of SAS consciousness. Following some commentary on this problem by Legrand (2007) and Legrand and Ruby (2009), I will examine the claims made in Rochat and Striano (2000) and Povinelli (2001), both of which are studies on the development of self-recognition in humans. I will devote considerable space to the discussion of Povinelli's work, as that work has important implications for my developmental study.

We have already encountered Legrand (2007) in the introduction. Recall her explanations of objects of experience as standing in opposition to the subjective experience that takes them as objects. What does the experiencing (the subject) cannot be equivalent to an object of one's experience. I have elaborated on this point at length in the *Conceptual* section.

Legrand claims that thus far, the results of experiments devised to investigate self-consciousness have been subject to flawed interpretations, owing to misconceptions of self-consciousness (2007). In particular, she criticizes those that equate self-consciousness with "identification and attribution to oneself of self-specific contents," (p. 593), that is, that equate self-consciousness with SAO consciousness. As Legrand and Ruby (2009) point out, such experiments fail to consider what it is about the contents that make them self-specific, namely that they refer to SAS (p. 273). I will consider a number of the cases that Legrand has criticized—and her criticisms of them—later in this section. Legrand and Ruby propose that all forms of self (i.e., SAO and bare SAS) must rely at least in part on self-specific processes (i.e., the mechanism behind SAS consciousness). They have yet to test this proposal empirically (p. 279), but their claim is supported by the results of the studies reviewed earlier in this section.

Rochat and Striano (2000) claim that the rooting reflex in newborn infants demonstrates an ability to attribute actions to self. Briefly, the rooting reflex is the orienting of the head toward

the side where the infant feels something brushing against the corner of her mouth (which is often a source of nourishment, e.g., the mother's breast or a bottle). Rochat and Hespos (1997) demonstrated that infants as young as 24 hours displayed the reflex three times as often when the corners of their mouths were touched by another than when they accidentally touched their own mouths. Rochat and Striano (2000) interpret this result as meaning that newborn infants can distinguish self and other at a rudimentary level. They claim that the self-touch is fundamentally self-specifying because it involves proprioception, which relates information only about the infant's own body. On this account, self-attribution is based on the integration of redundant proprioceptive and perceptual feedback; if both the sense of touching and sense of being touched occur simultaneously, the infant recognizes this as a self-touch and ignores it (p. 516). If their interpretation is correct, this would mean that infants become self-conscious at least as early as two months, if not sooner.

It is not clear, however, that Rochat and Striano's interpretation is accurate. As Legrand notes, the data leave unresolved the question of how this sensory redundancy translates into self-attribution. Rochat and Striano leave this question unresolved because they assume that self-consciousness consists merely of awareness of particular objects (Legrand, 2007, pp. 595-6). The rooting reflex merely shows that the child has learned to react differently to multiple sensations occurring at once than to a single sensation. There is no reason to suppose that this requires anything like SAS consciousness. To be fair, Rochat and Striano characterize the kind of self-knowledge available in early infancy as "implicit" self-knowledge (p. 514), suggesting a non-conscious process at work. This work has little bearing on *consciousness* of self, however, other than suggesting a possible base from which SAS consciousness might develop.

More pertinent to my investigation into how SAS consciousness develops is the work of Povinelli and colleagues on self-recognition in young children between the ages of two and five, summarized in Povinelli (2001). Povinelli's work is important for two reasons. First, it serves as another demonstration of the hazards of failing to recognize the SAS/SAO distinction. Second, while Povinelli's interpretation of his results is problematic, the results themselves are indicative of a key development in self-awareness, namely the integration of SAS and SAO consciousness.

The aim of Povinelli's research is to determine which aspects of the self are represented at various stages of early childhood. His central hypothesis is that, by the age of two, children possess a sense of self that is restricted to the present and immediate past and only later develop a sense of self extended into the past and future. He bases this hypothesis on observations made in earlier studies that children become capable of recognizing themselves in mirrors by the age of 18-24 months (Amsterdam, 1972; Lewis and Brooks-Gunn, 1979; cited in Povinelli, 2001). Povinelli's results are relevant to my developmental study, because the distinction I make between SAS and SAO suggests an alternate, more plausible interpretation of his results.

First, Povinelli and colleagues (Povinelli, Landau, & Perilloux, 1996) tested the hypothesis that the initial appearance of the self-concept (i.e., SAO, although he does not identify it as such) in human development was temporally restricted, effectively confined to the present. In an early experiment, children from the age of two to four years had stickers covertly placed on their heads during a play session. Each child was then shown a video three minutes later in which the experimenter could be seen placing the mark on the child's head. Povinelli expected that the child would reach up to remove the sticker upon realizing that (a) the events taking place in the video had just happened, and (b) that the child in the video was her.

Surprisingly, none of the two-year-olds and only a quarter of the three-year-olds reached up to remove the sticker. On the other hand, the majority of four-year-olds removed the sticker almost immediately upon seeing the sticker-placement event in the video. Informal questioning revealed that the younger children had some understanding that the child depicted in the video was herself. In several instances, upon being asked "Who is that?" while the experimenter pointed to the child in the video, the participant used the first-person pronoun (e.g., "that's me") or her proper name. There was no evidence to suggest that those who used the first-person pronoun were using it as a proper name and not as an indexical. When these same children were explicitly asked in a number of ways if they could get the sticker, however, they still failed to reach for the sticker. Povinelli (2001) remarks that they seemed to recognize the features of the image but failed to understand what these images had to do with them. On his interpretation, the child fails to understand the temporal relation between herself and the image. I suggest an alternative interpretation below.

In a follow-up experiment, Povinelli *et al*. (1996) sought to test whether the results of the experiment just described could be generalized, or that there was something about video images in particular that caused the surprising results. He performed a similar experiment using photographs of past events instead of video recordings, and this time his participant group was restricted to three- to four-year-olds. One photograph was taken when the experimenter placed the sticker on the child's head, and another at the conclusion of the game. After two to three minutes, the child was shown the pictures and asked the same type of questions as in the previous experiment. Those who did not reach up to remove the sticker were shown a mirror. The results were similar to those generated by the first experiment, with very few of the three-year-olds reaching up. Most of the children who did not reach up when shown the photograph

did reach up when shown the mirror. Once again, most of the younger children correctly identified themselves in the images using the first-person pronoun or their names, or by pointing to themselves, though the authors noted that use of the proper name was much more common among the younger children than the older ones, who used the personal pronoun almost exclusively. Furthermore, when children were asked questions about the events in the images, nearly all those who had reached for the sticker had a tendency to use phrases that established an association between themselves and the event. When describing the events, for instance, they would say things like "he's putting the sticker on my head." Those who did not reach for the sticker used dissociative phrases like "he's putting the sticker on *her* head" as often as they used the associative phrases.

It is not clear how to interpret this result, since some of the children who failed to reach up did use associative phrases. Povinelli (2001) claims that the younger children failed to recognize themselves in past images and videos because they failed to understand the temporal relations between past events and present states. However, there are more plausible interpretations of the results, on which I will elaborate later in this section.

To test his further hypothesis that children fail the task because they fail to understand the temporal relations between events and states, Povinelli *et al.* (1996) performed an experiment to see whether or not two-and-a-half- to three-and-a-half-year-old children responded differently to live versus delayed video feedback. Once again, a sticker was used to mark the children, and they were tested on whether or not they reached up for the sticker. In this experiment, children in the live feedback condition were shown live video while children in the delayed test group were shown video from three minutes earlier. The results were consistent with his expectations, with significantly more of the children reaching up during the live presentation than during the

delayed one. The overwhelming majority of children who did reach up in response to the delayed video only did so during the question period following the video.

These experiments suggested that a critical development in the child's ability to understand the events depicted in the videos and images occurred around the age of three. To more precisely determine the timing of this development, Povinelli and Simon (1998) devised another series of tasks for three- to five-year-olds. In one experiment, the experimenter covertly placed a sticker on the child's head during two different play sessions spaced a week apart. The child played a different game during each session. At the end of the second session, half of the children were shown a video recorded just five minutes earlier, while the other half were shown a video from a week prior. A large percentage of the four- to five-year-olds reached up for the sticker after the brief delay but not after the extended delay, while only about half of the younger children (from three to three-and-a-half years) reached up for the sticker after both the brief and extended delays (Povinelli & Simon, 1998).[11] Based on these results, Povinelli and Simon concluded that the ability to pass both immediate and delayed mark tests was not related to "featural recognition of self," (p. 193) an aspect of what I have called SAO consciousness.

Interestingly, Povinelli (2001) resists describing the young child's reaction to a mirror or other image of oneself as "self-recognition," claiming that this term is too specific. Instead, he prefers to describe the reaction as the child establishing an equivalence relation between the image in the mirror and the child's representation of herself. He claims that this explains why two- and three-year-old children can "recognize" themselves in mirrors yet have difficulty doing so in photographs and delayed video. Introducing a temporal delay breaks the equivalence relation, and the children no longer check for the mark (pp. 85-6). Note that on Povinelli's

[11] In the extreme-delay condition, the sticker was surreptitiously removed by the experimenter before the end of the session

account, consciousness of SAO is necessary for this equivalence relation to do any work in associating the image in the mirror with oneself. What is lacking is not the child's consciousness of herself as an object *per se*, but consciousness that the object *in the mirror* is her. Consciousness of SAO alone would not be sufficient to establish the latter; the child must also identify with SAS. Unfortunately, Povinelli does not recognize the SAS/SAO distinction, so his analysis is vulnerable to the infinite regress problem we encountered earlier.

The notion of self that Povinelli has in mind is made clear via his model of cognitive development. His model holds that in the second and third year, infants have access to a self-concept consisting of information about the infant's present kinaesthetic and mental states, which he simply refers to as the "present self." Now while children at the age of two and three are capable of retaining information about past states, Povinelli's model maintains that they have not yet developed the ability to represent these memories as temporally connected to representations of the present self (p. 83). On his interpretation, the crucial development is the ability to extend one's self-concept to include representations of kinaesthetic and mental states from previous instances in time. He speculates that this ability is actually a particular use of a domain-general ability to hold in mind "multiple, and contradictory representations of the same object or event," one's self-representations included (p. 83). This ability enables one to generate the concept of "self," consisting of "certain kinds of information about the self that were previously implicitly available," now rendered explicit (pp. 82-3) This information consists primarily of kinaesthetic information, which Povinelli maintains is "the most salient and omnipresent" for infants (p. 83). Notice that he characterizes the self strictly in terms of representations of particular objects.

Recall that some of the younger participants were able to identify the child in the delayed video but did not react to the appearance of the mark. Povinelli takes this to mean that children in

that age range have a sense of self restricted to the present. To explain how children are able to correctly name the child in the video, he speculates that earlier exposure to mirrors and the variability between which features of the images children attend to may cause them to establish equivalence relations with particular features of the images that change little over time, and are part of their current sense of self. On the other hand, the kinaesthetic information, which is held as primary, informs the participant that the image cannot be equivalent to her (p. 86). Thus on Povinelli's account, the participant is torn between conflicting representations, causing her to produce the paradoxical response observed. The entire explanation, unfortunately, is a speculative leap and not supported by the results.

Povinelli takes the results of the latter series of experiments to mean that different psychological processes are responsible for the child's reaction at different stages of development. Specifically, younger children who understand the equivalence between themselves and the video image do not yet understand the time difference in the video, so they reach for the sticker regardless. Older children also understand the time difference, and only reach for the sticker after the brief delay, seemingly because they reason that if the event just occurred, the sticker will still be on their head. If the event occurred during the last visit, they would reason, it should not still be on their head. Povinelli's results show that the response to the brief delay increases with age while the response to the extreme delay decreases, which is consistent with this explanation (pp. 89-90).

The interpretation Povinelli provides, however, is problematic because it is based upon a model of the self defined entirely in terms of particular representations. He appears to have come close to the SAS/SAO distinction in identifying that a full sense of self requires a common point of reference among one's representations. His error is to characterize this point of reference as

the representation of a particular object, albeit extended in time. Recall Shoemaker's (1968) argument that the self cannot be identified with any particular object, because to identify a particular object with oneself would require one to identify something true of SAO that is also true of the self.[12] Since Povinelli does not recognize the SAS/SAO distinction, he supposes that any confusion with respect to identity of the physical body is evidence for a lack of understanding with respect to the persistence of self across time.

Povinelli's work does not serve merely as an example of a failure to recognize the SAS/SAO distinction—while his interpretation is suspect, his results point to an interesting development in the child's understanding of self, one that can better help us understand the mechanism of self-awareness, specifically the interplay between SAS and SAO consciousness, and how these are related to key cognitive faculties like episodic memory. Povinelli did not test participants' capacity for episodic memory, which children in the age groups he tested would be expected to have. The possession of episodic memory, specifically that of the autobiographical variety (episodic autobiographical memory, or EAM), would count against Povinelli's interpretation because it would demonstrate that children have an understanding of self extending at least into the past. In my developmental study, described in *Empirical Studies*, I measure both EAM and delayed self-recognition capacity to test Povinelli's interpretation.

Povinelli claims that the crucial change, occurring roughly between the ages of four and five years, that allows children to understand previous instances of themselves *as such* is a basic

---

[12] As I mentioned in the *Conceptual* section, identifying my image in a mirror as *me*, for instance, requires that I recognize something true of the image that is also true of my body. Likewise, to recognize the bodily self-image as my own requires that it share some feature with my self-concept, which may include, for example, spatial information. At some point, the "object" held in common must be the featureless self (i.e., SAS), because identifying a feature-laden object with the self would require me to posit something further to be held in common. So long as the feature held in common is itself an object with features, we will need to posit further entities, leading to an infinite regress (Shoemaker, 1968, p. 87). Since our minds are free of such regress, we must posit the featureless SAS.

understanding of causality. That is, they possess a sense of self rooted in the present but fail to understand how causality works. While I agree that such an understanding plays an important role in this development, I propose that the key development is the integration of consciousness of SAS and SAO. Povinelli fails to account for this possibility because his account relies on a notion of self restricted to representation of particular objects. SAS consciousness enables one to recognize that multiple experiences share a single common subject, and in turn enables one to recognize previous instances of SAO as belonging to the subject, but at a different point in time. This recognition is best illustrated by the capacity for EAM, which—as I will argue in the first section of Chapter II, and explore via my studies in the final section—requires SAS consciousness.

I claim to provide a better interpretation of Povinelli's results by taking the SAS/SAO distinction into account. My claim is *not* that the crucial change is the onset of SAS consciousness. Indeed, children who failed the delayed self-recognition task had no difficulty attributing objects of representation to self. My claim is that integration of consciousness of SAS with consciousness of SAO comes in degrees, beginning with a sense of self extended over a period limited by working memory capacity. This explains why it is limited to current experience—a delay period of only three minutes, on average, was enough to prevent the younger children from reaching for the sticker. At this stage, a child may be able to recall previous instances of SAO, yet not recognize them as belonging to self. As the child develops, she is able to recall previous instances of self, and recognize them *as* self, because she recognizes the common element, SAS, in both. That is, she recognizes the memory of a past self for what it is. Prior to this stage, she may well have memory of *someone*, with a particular experience, that she can correctly label with a proper name. What she lacks is an integrated sense of self, the

combination of SAS with past instances of SAO. As SAS consciousness develops, the child begins to retrieve information about previous instances of SAO from long-term memory, integrating them into a temporally-extended sense of self. Like many developmental processes, the ability to integrate consciousness of SAS with consciousness of SAO is likely a gradually-acquired capacity, one that the child becomes better at reconciling over time, making fewer and fewer misrecognitions over time.

On my interpretation, the child is able to associate subject with object when the feedback is immediate (e.g., presented via mirror or live video), but fails to do so when a delay is introduced. In other words, the ability to identify self as the common subject of experience may be restricted to only a small subset of remembered experiences rather than all of them, as I would expect from an older child. The four- or five-year-old's fully developed SAS consciousness is what enables her to reason that the sticker is still on her head. The change in ability to recognize the difference between the brief and extreme delay could be explained as an increased ability to make the connection between subject and object over longer periods. To determine whether or not this interpretation is accurate, I need to further investigate the relationship between episodic memory and self-reference. As I will discuss in Chapter II, existing studies have only begun to explore the self-referential aspects of episodic memory.

There are also methodological reasons to suspect Povinelli's delayed self-recognition paradigm as a measure of temporally-extended self-awareness. Lind and Bowler (2009) found that children with autism spectrum disorder performed comparably to developmentally-normal children of the same mental age when given the delayed self-recognition task. However, the autistic group performed poorly on theory-of-mind tasks, and had difficulty with use of the first-person pronoun. If delayed self-recognition were a reliable measure of temporally-extended self-

awareness, we would expect that children lacking it—as indicated by poor theory-of-mind performance (as measured by the "Smarties" false-belief task, originally used by Perner, Frith, Leslie, & Leekam, 1989) and first-person pronoun use—would show correspondingly poor delayed self-recognition ability. Theory-of-mind performance was measured using Perner *et al.*'s (1989) "Smarties" false-belief task, where participants are asked about their own and others' false-beliefs about the contents of a box labelled as Smarties.

Earlier studies demonstrated that children with autism with a mental age of 18 months could pass the mirror recognition task, showing at least awareness of the present self (Dawson & McKissick, 1984; Ferrari & Matthews, 1983; Neuman & Hill, 1978; Spiker & Ricks, 1984). Lind and Bowler expected, however, that children with autism would have difficulty understanding self extended in time because that would require some metarepresentational ability that children with autism lack, such as the ability to hold multiple representations of self in mind for the purposes of comparison.

Lind and Bowler (2009) suggest two possible interpretations of their curious results: either children with autism have no deficiency in temporally-extended self-awareness, or the delayed self-recognition task is not a reliable measure of temporally-extended self-awareness (p. 648). Lind and Bowler favour the latter, because, as Povinelli (2001) suggested, it is possible to perform the task without relying on a temporally extended sense of self. Povinelli even admits that a small minority of children may have focused on the similarities between the delayed images and the present self-concept—enough to establish an equivalence relation—thereby passing the test without understanding that the image is from the past (p. 86).

Another reason that Lind and Bowler suspect the delayed self-recognition paradigm is, as earlier studies demonstrated, that children with autism have impaired episodic memory (Boucher

& Bowler, 2008; Bowler *et al*. 2007). This limitation suggests some deficiency in temporally-extended self-awareness. Furthermore, Lind and Bowler argue that *even if* delayed self-recognition is a reliable measure of temporally-extended self-awareness, delayed self-recognition would only be informative about one's sense of physical continuity. It would tell us nothing about one's sense of mental continuity or SAS consciousness. They suggest that future studies should focus on mental continuity, specifically that provided by episodic memory (p. 648-9). I will pursue this link between self-awareness and episodic memory later in the work.

Summary

I have reviewed work that either mentions or presupposes the SAS/SAO distinction, and that which fails to recognize the distinction but would benefit from taking it into consideration. The work on misattribution of other's features and actions to self demonstrates that the distinction is meaningful, because it suggests a mismatch between our sense of self and the objects we attach to it. If the distinction were not meaningful—if the objects *just are* the self—it would be difficult, if not impossible, to make sense of misattribution errors. There can be no mismatch where there are not at least two things to compare.

On the other hand, researchers who have attempted to account for attributions or misattributions to self without recognizing the SAS/SAO distinction have had difficulty generating plausible interpretations of their results. Rochat and Striano's account could not specify what made some contents self-specific rather than others. Povinelli's account came close to recognizing the distinction, but posited the need for a specific representation as the element common to experience. It is unclear what could serve as this common element. Furthermore, his explanation is that children lack a sense of a self persisting over time. Research on episodic memory, which I describe in Chapter II, suggests that this explanation is highly unlikely.

Children do seem to have a sense of a persistent self, specifically SAS. I claim that what they lack is not a general ability to understand the self as temporally-extended, but a specific difficulty associating past instances of SAO with self. In my developmental study, described in the final section of *Empirical Studies* in Chapter II, I explored this claim.

In the remainder of this work, I take the SAS/SAO distinction as a basic assumption. The existing conceptual and empirical evidence that I have reviewed is sufficient to justify doing so, and I will not provide additional empirical evidence in support of it. In the first half of Chapter II, I examine empirical work that supports the idea that SAS consciousness has the three distinctive features that I identified in the *Conceptual* section—namely, indexical self-reference, universal availability, and non-attributive self-reference. I will review what little is currently understood about the mechanisms underlying SAS consciousness, and indicate what I think are the most promising directions for further empirical research. In *Empirical Studies* in the following chapter, I will follow up on these directions with a pair of studies I conducted, one complete and one preliminary.

Chapter II – Studies

Introduction

Having completed the conceptual groundwork and illustrated the need for the SAS/SAO distinction in Chapter I, Chapter II will describe a pair of studies undertaken to investigate the nature of SAS consciousness. In the first section, I give a selected review of empirical literature on cognitive faculties that indicate the presence or absence of SAS consciousness. I have selected works that provide evidence for the features of SAS consciousness that I expounded upon in the *Conceptual* section, namely indexical self-reference, universal availability, and non-attributive self-reference. These works provide important background information for my studies. For the purposes of the empirical review and subsequent studies, I will be treating universal availability and non-attributive self-reference as the same phenomenon. Due to their highly-interleaved nature, existing empirical methods are still too coarse-grained to tease them apart, and there is no existing research that explicitly identifies either phenomenon.

In the final section, I provide a detailed overview of two original studies. The first study looks closely at the relationship between episodic memory and SAS consciousness, and focuses on the effects of inattention to self on SAS consciousness in adult participants. The aim of this study is to explore how the formation of episodic memories may be dependent upon SAS consciousness, such that the attenuation of the latter would result in a decrease in episodic memory formation.

The second study is a preliminary investigation into the development of self-awareness in early childhood, specifically on the integration of SAS and SAO consciousness and the relationship between episodic memory and SAS consciousness. This was originally intended to

be a full study, however there were numerous difficulties recruiting enough children in a timely manner, as I discuss in *Empirical Studies*.

The aim of this final investigation is twofold. First, I claim that by the age of three-and-a-half years, typically-developing children are capable of SAS consciousness. I explore whether or not children display evidence of the key features of SAS consciousness mentioned above. Second, I claim that before the age of approximately five years, children have yet to fully integrate SAS and SAO into a coherent sense of self. I claim that this integration plays a key role in the development of episodic memory, and I hope to provide some insight into the nature of that role.

Research on Indicators of SAS Consciousness

Introduction

Recall that one of the hallmarks of SAS is indexical reference. Though we can indexically self-refer without language (such as pointing to oneself), *correct* use of the first-person pronoun is a good indicator of this ability. In the first part of this section, I will examine the psycholinguistic work that has been done on use of the first-person pronoun. The literature on mastery of the first-person pronoun (Blum, 2005; Clark, 1978; Cole, Oshima-Takane & Yaremko, 1994; Oshima-Takane, 1985 & 1992; Oshima-Takane & Oram 1991; Schiff-Myers 1983)—where 'mastery' implies use as an indexical rather than a proper name—is surprisingly scant and methodologically suspect, and I include it largely as a further indicator as to where there is work left to do on the development of self-awareness.

I will also consider empirical research pertaining to the remaining two features of SAS consciousness covered in the *Conceptual* section, the universal *availability* of SAS and the related non-ascriptive self-reference. As I argued in that section, SAS is part of the *structure* of

experience. Being part of the structure, and therefore not dependent on particular contents, it is always available in any experience—one can reflect on it at any moment. Being independent of particular contents means that SAS can also be referred to without ascribing (i.e., attributing) any content to self.

Following Brook (2001), I have claimed that, being content-less, SAS does not divide experience. This means that I cannot compare one part of SAS with another part (it has no parts) or compare the present SAS with earlier instances, because there are no features that can differ. When I refer to SAS, it will appear as the same subject that I referred to yesterday, last week, or ten years ago, and will appear as the same subject that I refer to in future moments. SAS is the common tie that binds experience across time. If we want to study SAS consciousness, then, one place to start is to look at those faculties involved in binding experience across time. The most obvious starting point is autobiographical memory—our memory of key moments in our personal history. However, even past events without personal significance can be recalled as events that *I* experienced, that is, recalled from the standpoint of having the experience.

The term "autobiographical memory" picks out a system of memory organized around specific contents – those events that hold personal significance, but this system of memory is dependent upon the contributions of at least two other memory systems, namely episodic and semantic memory. The episodic components of autobiographical memory (hereafter, what I refer to as episodic autobiographical memory or EAM) are of most interest to the current project.

Episodic memory is, roughly, the encoding and recall of experienced events, in particular the *what*, *where*, and *when* of the events in question (Tulving, 2002, p. 3). Episodic memory is a subset of declarative memory, and is distinct from the other subset, semantic memory. By way of

contrast, semantic memories are generally thought of as memories of facts, such as ‘Paris is the capital of France.’

Indeed, the colloquial use of the terms *memory* and *remember* mostly capture the notion of episodic memory, while the terms *knowledge* and *know* are used for semantic and procedural memories (Tulving, 1985). For example, I *remember* my sixth birthday, but I *know* that mangoes grow on trees and I *know* how to ride a bicycle. I have not personally experienced a mango tree, however, and although I know that I learned how to ride a bicycle, I cannot re-experience the event mentally.

This usage is the basis for a popular experimental paradigm in the study of episodic memory, the *remember/know* paradigm (e.g., Tulving, 1993; Tulving & Kroll, 1995), where a participant’s ability to recall episodic details is keyed to use of the terms ‘remember’ and ‘know’. I am more likely to report, for example, that I *know* that Kylie recently celebrated her 30th birthday if I did not personally attend her party but was aware of it, whereas I will likely say that I *remember* her birthday if I did attend. In the remember/know paradigm, participants give confidence judgments on a three-point scale—‘3’ being the highest—about whether or not a word appeared in a list. If they remember the word being in the list—that is, if they can recall it episodically—they are instructed to give a rating of ‘3’, whereas if they know that the word was in the list but cannot remember seeing it, they are to give a rating of ‘2’, and if they are merely guessing that the word was in the list, they give a rating of ‘1’.

It is interesting that a fundamental difference between the two types of recall lies in the relation between *me* and the events in question, that is, whether or not the event was *experienced by* me. Self-awareness is thought to be a crucial component of episodic memory (Tulving, 2002, p. 2-3; Nelson & Fivush, 2004). Along these lines, I claim that episodic memory is dependent

upon SAS consciousness. A consequence is that in the absence of SAS consciousness, episodic memories would not be able to form.

Under circumstances where one is frequently distracted from SAS, I expect to see a decrease in ability to encode episodic memories. The first study in *Empirical Studies* tests this claim. To my knowledge, no previous research has examined the relationship between SAS consciousness and attention. However, Gardiner (2001) and Gardiner, Gregg, and Karayianni (2006) have done research on the relationship between episodic memory (which I claim is dependent upon SAS consciousness) and attention which provides me with a starting-point for my own research. I will also review literature on the effects of distraction on visual attention, specifically cases of inattentional blindness (Vetter, Butterworth & Bahrami, 2008), as I hypothesize that a similar mechanism applies to attention directed at SAS. Carver and Scheier (1978) found that exposure to a mirror causes an *increase* in self-related thoughts. Participants were asked to complete sentence stems, and those who had been exposed to a mirror were more likely to complete sentence stems with self-referential language and no reference to the external world. This result indicates that self-directed attention can be manipulated experimentally, with the caveat that 'self-directed attention' was defined in very general terms. I describe Carver and Scheier's results in detail in *Effects of Inattention on Episodic Memory and SAS Consciousness*, below.

I also claim that children lack a fully integrated sense of self before at least the age of four years, as I indicated in my interpretation of Povinelli's findings. If so, I expect to see corresponding deficiencies in episodic memory. There is some debate in the literature over when episodic memory develops. I review two studies that differ widely on this point. Vandekerchove

(2009) places the development of episodic memory as late as the age of four, while Peterson (2002) claims that children can encode episodic memory as early as one or two years.

It is important not to overstate the connection between episodic memory and SAS consciousness. While episodic memory may be dependent upon SAS consciousness, it is possible to have SAS consciousness in the absence of episodic memory. For example, Piolino *et al.* (2009) shows that patients with EAM (episodic autobiographical memory) disruptions are still aware of SAS (they can still ask "who am I?"), even though they are unaware of themselves as objects extended into the past. Similarly, Addis and Tippet (2004) demonstrated that in Alzheimer's patients, loss of EAM leads to substantial changes to sense of identity (i.e., SAO) but there is no indication that these patients fail to understand themselves as subjects of experience (SAS).

The case studies of H.M. (Henry Molaison) and K.C. (for a review, see Tulving, 2002) also demonstrate that one can be conscious of SAS even in the absence of episodic memory, or indeed any long term memory whatsoever. The case of K.C. demonstrates the former, while H.M. demonstrates the latter. As a result of a motorcycle accident, K.C. suffered extensive bilateral damage to much of his hippocampus and medial temporal lobes, such that he could not form any new memories. While he could remember facts about his past from the time before the accident, he could not recollect personal events episodically, nor could he imagine himself in future scenarios. Despite these deficiencies, however, K.C. shows every indication that he has awareness of self in the present (Rosenbaum, Köhler, Schacter, Moscovitch, Westmacott, Black, Gao, & Tulving, 2005). H.M., on the other hand, was unable to form long-term memories (with the exception of procedural memories—he could learn how to do new things, even though he

could not recall learning them) yet showed no indications that he failed to be aware of himself as the subject of his experiences—he had no trouble using the first-person pronoun, for example.

Indexical Self-Reference

As I argued in the *Conceptual* section, one of the hallmarks of SAS consciousness is the ability to self-refer indexically. I expect, then, that a good indicator of SAS consciousness is mastery of the first-person pronoun. The reasoning is that in order to properly use the pronoun "I," I must understand that it refers to self. In order to use the pronoun correctly, a child needs to come to the realization that "I" does not single out a specific object in the world (in the way that "tree" or "Sarah" does), but is relative to the speaker. This in turn requires at least some vague notion that self and others are *subjects* of experience.

No research has attempted to address the specifically indexical nature of first-person pronoun use, although an extensive amount of research has been done to ascertain *when* children master the first-person pronoun. In my developmental study, I measured childrens' mastery of the first-person pronoun, using a positive result as an indicator of SAS consciousness.

Evidence from psycholinguistic studies places the first use of the first-person pronoun at around 20 months (Blum, 2005, cited in Meissner, 2008), with mastery occurring some time later, anywhere from 25 to 34 months, with significant variation across studies (Oshima-Takane & Oram, 1991; Oshima-Takane, 1992; Schiff-Myers, 1983). In children with hearing impairments, there is a slight delay in mastery of pronoun use, but this delay is likely due to a lack of linguistic experience (Cole, Oshima-Takane, & Yaremko, 1994). Mastery is determined as the time at which the child no longer confuses "I" with "you." Clark (1978) and Oshima-Takane (1985) show that a minority of children continue to demonstrate some confusion about the use of pronouns past the age of three. These results are interesting because they place the

time of mastery of the first-person pronoun in the same time period or slightly before children develop the ability to recognize themselves in videos and photographs (see Povinelli, 2001).

There are a number of methodological difficulties in measuring correct pronoun use, however, as Lee, Hobson, and Chiat (1994) point out:

> [S]poradic use of pronouns often precedes more systematic and frequent use… it is often difficult to determine whether the pronouns have achieved syntactic independence from the phrases in which they are embedded, and… errors may reflect specialized rather than deficient forms of pronoun usage... (p. 159)

This quote also serves to emphasize the importance of measuring correct pronoun use as opposed to simply measuring how often the child uses the pronoun. Although Lee *et al.* did not look at time of mastery in typically-developing children, their study is of interest because of the careful design of their tasks. The authors were particularly interested in abnormal use of pronouns by children with autism—motivated in part by the postulate that children with autism are deficient in sense of self as agent and owner (see Bosch, 1970; Silberg, 1978)—so it was especially important to control for the cited difficulties. They used developmentally-delayed children as a control group (the authors do not specify the type of delay). The children (and some young adults) with autism ranged in chronological age from 8 to 23 years old, with verbal mental age ranging from 3 to 8 years old. Children in the control group ranged in chronological age from 8 to 25 years old, and ranged in verbal mental age from 3 to 8 years old.

The authors evaluated pronoun use (i.e., production) via three tasks. They also administered a parallel series of comprehension tasks which I will not elaborate on, as there were no significant differences between groups on those tasks. The first production task involved teaching the participant the names of pictures on double-sided cards, then holding up the cards and asking them to identify who sees the picture on the front or back. The idea was that children

who understand the correct use of the first-person pronoun should say ‘me’ or ‘I’ when asked to identify the picture facing them, and ‘you’ when the picture is on the opposite side. Each participant was given three pictures where the correct answer was “you,” and three where the correct answer was the first-person pronoun (‘I’ or ‘me’). If the participant gave the same answer on at least two out of three trials, that answer was recorded as the predominant response. The authors found no significant difference between groups. Most of the participants in each group answered correctly. Children with autism were found to use the pronoun “I” significantly more than the controls, who predominantly used ‘me.’ (pp. 160-3)

The second task measured which pronouns were used to refer to pictures of the participant and experimenter. Participants were shown two slightly different pictures of the experimenter, and likewise of the participant. They were also shown two pictures of peers. The experimenter determined a week earlier that the participant knew the names of the people in the pictures, without using the pictures themselves. Participants were shown three pairs of photographs: the experimenter and the participant, of the participant or experimenter and the first peer, and of the second peer and the participant or experimenter. Half of the participants were shown the photo of the experimenter first, while the other half were shown the photo of the participant first. The participants were then asked, “Who is this a picture of?” Their answers were scored by placing them into one of three nominal categories: use of pronouns alone, use of pronouns and names, or use of names alone. There were no significant differences in pronoun use between “upper ability” (high linguistic ability) participants with and without autism, however there was a significant difference between the lower ability children with autism, 10 of 12 using names instead of pronouns versus 4 of 12 typically-developing children. Upper ability

children with autism used “me” significantly more than other cognates of the first-person pronoun.

In the third task, the experimenter told participants that they would be shown a series of pictures, and they would have to identify who was in each. The experimenter showed three pictures of peers, followed by a picture of the experimenter or participant, then another three peer pictures followed by a picture of the experimenter or participant. Half of the participants were shown the experimenter first and then the participant, and vice versa for the other half. As in the previous experiment, children’s responses were scored in terms of whether they used pronouns alone, pronouns and names, or names alone. There were no significant group differences in the use of the first-person pronoun, although lower ability children with autism were significantly more likely than upper ability children with autism to use proper names to refer to themselves rather than the first-person pronoun.

It should be noted that although mastery of the first-person pronoun *does* indicate SAS consciousness, the most that failure to master the pronoun demonstrates is that the child is lacking in linguistic ability, and does *not* indicate that a child lacks SAS consciousness. The import of this study for my purposes is the design of the tasks, which seem well suited to evaluate correct use of the first-person pronoun. In the *Empirical Studies* section, I use Lee *et al.*’s first picture task and a variant of the second picture task to evaluate first-person pronoun use.

Insofar as indexical reference is concerned, the above studies strongly support the claim that children have SAS consciousness at least by the age of two (in the third year), on average. Thus far, however, there has been no research on the special nature of first-person indexicality. I do not investigate the development of this feature in any more detail in my own study, as I do not

have access to a suitably young population. I did, however, test for use of the first-person pronoun among 3- to 5-year-olds, simply to provide additional evidence for SAS consciousness in those age groups.

The other features of SAS consciousness—universal availability and non-attributive self-reference—are easier to study in the age-groups I will be testing. My first study will expand on behavioural correlates of these features in adults while my second, preliminary study will explore the development of these features. In the following section, I review what existing research can tell us about universal availability and non-attributive self-reference, highlighting work with particular relevance to my own studies.

Universal Availability and Non-Attributive Self-Reference - Episodic Memory and SAS

In the *Conceptual* section, I argued that SAS is universally available across all experience, such that reference to SAS is non-attributive. Due to the close interconnections between universal availability and non-attributive self-reference, it is too difficult to extricate them for the purposes of empirical study. In order to separate universal availability and non-attributive self-reference, I would have to identify a behavioural correlate of a universally available state that did not involve non-attributive reference. Currently available research does not distinguish between these two features of self-awareness, so there is no precedent with which to divide studies. I will therefore treat the pair as a single phenomenon throughout the remainder of the discussion. I expected that the results of my studies would provide some means for identifying universal availability and non-attributive reference as distinct features.

To determine the psychological mechanisms of SAS consciousness, a reasonable starting point is to examine cognitive faculties that are universally available and that do not depend on specific attributions. The expectation is not that I will find a faculty that corresponds exactly with

SAS consciousness, but that I will find one that depends upon SAS consciousness, such that the absence or attenuation of that faculty would be indicative of the absence or attenuation of SAS consciousness.

Episodic memory is one faculty where the features of universal availability and non-attributive self-reference are clearly displayed, which suggests that SAS consciousness plays an important—and, as I will claim, foundational—role in episodic memory. Episodic memories are distinct from other types of memory in that they are remembered *as experienced by* the person who witnessed them. When I recall an episodic memory, I recall it from *my* standpoint. Along with the knowledge of *where* and *when* the event happened is the knowledge that it was I *who* witnessed it. Since the event is encoded with information about who experienced the event, episodic memories are necessarily self-referential, a point that I expand on below. Other types of memory *may* refer to self, but need not. Furthermore, we are aware of the subject of the recalled episode and the subject of the present instant as one and the same subject, regardless of how much time has passed since the remembered event. In this way, episodic memories tie earlier instances of self to the present, and may provide the basis for temporally-extended self-awareness, as Lind and Bowler (2009) suggested.

Episodic memories are distinct from semantic and procedural memories, which are memories (knowledge) of facts and skills, respectively. Semantic memories are recalled free of context; you can remember some piece of information without knowing where or when you learned it. Semantic memories may originate as episodic memories – for example, when you first learned that Paris was the capital of France you may have been able to recall a short time later where you were when you learned it and when you learned it. Over time the most relevant details

of that memory are encoded into long-term memory as semantic memories, that is, memories of facts divorced from context.

Procedural memories are largely unconscious, and provide you with the knowledge of *how* to do something (Wheeler, Stuss, and Tulving, 1997). An often-used example of procedural memory is the knowledge one has of how to ride a bicycle. You can learn how to ride a bicycle at a very young age, yet it is extraordinarily difficult to describe to someone how to ride a bicycle, because the knowledge is largely unconscious.

Episodic memories are thought to involve semantic memories, but encode the context – that is, the *where* and *when* – in addition to particular facts. I might remember that I first learned to ride a bicycle in the driveway of my parents' house when I was five, for example. Episodic memories are also unique in the manner in which they are recalled; episodic memories are recollected from the standpoint of the person having them.[13] As Wheeler *et al.* (1997) explain, "[e]pisodic recollection is well described, with reference to that personal feeling experienced when a rememberer reflects on some moment in the past." (p. 333).

Another purportedly key element, and one that is particularly important for the current project, is that episodic memories are recalled as events that happened *to*, or were *experienced by* the subject. As Nelson and Fivush (2004) explain, "Episodic memory… is always specific in terms of location of an event in both time and space, as well as in the specific awareness of self in the experience—the feeling that 'I was there, I did that.'" That is, in the terms I have been using, there is a necessary recognition of SAS in episodic recall. Wheeler *et al*. (1997) depict this relationship as one of necessity:

[13] This property of episodic memory is sometimes called *autonoetic consciousness* (Tulving, 2002). However, in the interests of terminological clarity, I avoid using that term here.

> [I]t is the awareness of self… that serves as a foundation for this unique capacity of human consciousness [referring to the sense of re-experience]. Only through the sophisticated representation of self can an individual autonoetically [see footnote on p. 78] recollect personal events from the past and mentally project one's existence into the subjective future. (p. 334)

The statement that self-awareness (at least SAS consciousness, though the authors fail to make the SAS/SAO distinction explicit) is foundational with respect to episodic memory is the keystone of the studies that I describe below. This connection between self-awareness and episodic memory is almost universally acknowledged, yet there has been very little research devoted to it (Nelson & Fivush, 2004). I will review what little has been written on this connection in the following section.

The definition of episodic memory does not imply that such memories are eidetic (recalled in "photographic" detail), or even recalled with great accuracy. It is understood that the re-experiencing is also a reconstructing of events, and highly susceptible to the influence of the person's imagination. Indeed, in the legal system, eye-witness testimony is considered unreliable because of this fact. As Nelson (1993) notes, "[a]lthough the validity of a memory may be of concern if one is interested in such issues as whether or not children are reliable witnesses, it is of less concern if one is interested in when they begin to retain memories… Memories do not need to be true or correct to be part of [a memory] system." (p. 8)

What is important for my purposes is that, accurate or not, episodic memories are experienced in the same manner in which we experience the present, as events witnessed by a common subject (that is, SAS). Many researchers (Atance, 2008; Tulving, 2002) describe the manner of recall as "mental time travel" into the past, as the experience of episodic memory is

akin to "playing back" previous events.[14] It can be thought of as playing through a sequence of events in the same order in which they occurred (though it is possible to be wrong about the order). It is best thought of in contrast with semantic memories, which are largely propositional in nature. Episodic memories can be described propositionally, but are experienced in much the same way present events are experienced, that is, from the standpoint of the experiencer.

An important distinction must be made between episodic memory and autobiographical memory, as the differences are often unclear in the literature. The difficulty is that procedural, semantic, and episodic memories are distinguished by the manner in which they are presented in consciousness, whereas autobiographical memories are distinguished from other kinds of memory by their *contents*. Nelson and Fivush (2004) conceive of autobiographical memory as a *functionally* distinct memory system, in contrast with the *neurologically* distinct episodic memory system. They define autobiographical memories as memories of events that are particularly significant for one's sense of personal identity, that is, that contribute to one's sense of a coherent self extended into the past. Conway and Rubin (1993) observed that autobiographical memory is closely linked to the reminiscence bump, the observation that people have far more memories of events that occurred during periods of formation of personal and political identity than any other life periods.

Further complicating the distinction is the fact that most—but importantly, not all—autobiographical memories are also episodic, although the reverse is not true. For example, certain memories of my wedding day have great personal significance and can be remembered episodically, whereas I can remember watching TV last night, though the event has no personal meaning to me.

---

[14] I caution that we should not take the electronic recording metaphor too literally, as past events are recalled with far less fidelity than the metaphor suggests. The term "mental time-travel" is problematic for the same reason.

As I mentioned, autobiographical memories are not always episodic (Conway, 2005), though they are often treated as if they are (e.g., Nelson & Fivush, 2004). Autobiographical memories are complex constructions that combine an autobiographical knowledge base—a store of self-related semantic memories, or facts about self—with knowledge of general events (e.g., birthdays, weddings, funerals) or episodic memories of specific events (e.g., your 30th birthday, your cousin's wedding, or your grandfather's funeral). A person can retain certain autobiographical memories, such as the fact that she enjoys playing the piano, or moved to the city when she was twelve, even when episodic memory is impaired or otherwise inaccessible. For example, Rathbone, Moulin and Conway (2009) document a case of retrograde amnesia where the patient—P.J.M.—was able to maintain autobiographical memories even though she lost specific *episodic* memories of the 18 months leading up her brain injury. When asked to recall memories from this period of her life, she would typically generate facts about herself but not memories. It is important to note that she could generate *some* episodic details, such as what she felt when her boyfriend told her he was going to South Africa; this capacity was greatly diminished, however, relative to participants in an all-female control group. P.J.M. could construct narratives by combining generic events—for example, joining a club, getting married, or having a child—with semantic facts about herself, and she could infer from a given fact that some event must have taken place (e.g., from the fact that she moved to a new house, she inferred that she must have moved because there wasn't enough space in the old house, even though she could not remember choosing to move). P.J.M is not the only documented case of maintaining a sense of self in the absence of episodic memory. Tulving (1993) reports that K.C. knows semantic facts about his own traits, yet cannot recall any events illustrative of those traits.

Autobiographical memories need not be episodic, but I will focus on those that are—what Nelson and Fivush call episodic autobiographical memory, or EAM. It is the episodic aspect of such memories—being remembered *from the standpoint of* the person having the experience—that requires SAS consciousness. It may be that I can maintain a sense of self without access to EAM, using semantic memory—for example, knowing *that* I attended my 30th birthday party—even if I cannot recall the details of the event episodically. While it is *possible* for semantic memories to be self-referential, however, it is not *necessary*. Episodic memories, on the other hand, because they are encoded in the manner in which they are experienced—from the standpoint of the experiencer, with details recalled in the same chronological order as they were witnessed—are necessarily self-referential. The differences between episodic *non*-autobiographical memory and EAM are principally in terms of content, although merely-episodic memories are typically forgotten after as few as 24 hours (Conway, 2005, p. 613). For my purposes the distinction is relatively unimportant, since self-reference is common to both types of episodic memory. The work on episodic memory is significant for my project, because the faculty of episodic memory appears to require SAS consciousness.

As I have mentioned, there is a link between self-consciousness and episodic and autobiographical memory. These kinds of memories clearly play a role in one's sense of SAO, allowing a person to compare her present features with earlier ones, giving her a sense for how she has changed—or not changed, as may be the case—over various lengths of time. As Rosenbaum *et al.* (2005) observe in a case study of severe amnesiac K.C., the absence of episodic memory in particular can lead to profound changes in one's personality. Although K.C.

has limited access to some autobiographical facts,[15] he lacks the ability to recall them as having happened at a particular place and time, and as Rosenbaum *et al.* note, "any narratives that he managed to produce [with regard to locations or photographs relevant to his personal past] lacked the subjective re-evoking of the emotional and contextual details that define a personal from a non-personal episodic experience."

Recall my claim that one's sense of SAS, particularly one's sense of being the subject of both past and present experiences, is required to encode episodic memories. There are no obvious signs that K.C. lacks an understanding of himself as SAS, either in the present or across experience, and it is unlikely that a deficiency of SAS is responsible for his condition, so his case is of limited import to my studies. Developmental studies of episodic memory are of greater import because they provide a precedent for the methods I use in my own preliminary study of children, and reinforce my expectation that children between the ages of three and five years are capable of episodic memory and thus possess SAS consciousness.

There are a small number of studies investigating the development of episodic memory in young children. Vandekerckhove (2009) claims that prior to the age of four, children may be able to recall past events, but do not encode them in the manner that they were experienced (i.e., episodically). If her claim is correct, it would indicate that children don't recognize a common subject of experience extended into the past, as past experiences have the same status as impersonal information about the world. Her interpretation is that children in this age group lack the ability to re-experience events, or lack a sense of themselves "within subjective time and space" (p. 9). These children may have semantic memories of the details of past events, but no

[15] K.C. is considered deficient in this respect, as measured by a standard test of autobiographical memory, the Autobiographical Memory Inventory (Rosenbaum *et al.,* 2005; More information on the AMI test can be found in Kopelman, Wilson, & Baddeley, 1989 & 1990).

episodic memories. It is generally agreed—with some exceptions, discussed below—that children younger than four do not have a fully-functional episodic memory system (Perner & Ruffman, 1995; Tulving, 2002).[16]

Vandekerckhove (2009) claims that the absence of episodic memory explains Povinelli's finding that 2-year-olds could not recognize themselves when shown a briefly-delayed image. It is not until the age of 4 or 5 when children come to fully understand how the present image is related to past images. Children have semantic memory of the image, but don't remember the self-related, time-related context of it (p. 9). She also claims that children need to be able to reflect on past experiences in order to recall them episodically.

Counter-examples to the preceding account are given by other studies that purport to show that children can remember earlier events from as young as two-and-a-half years (Eacott & Crawley, 1998; Terr, 1988). The methods employed in these studies, however, leave an open question as to whether those children are relying on semantic or episodic memory (Vandekerckhove, 2009, p. 10).

Vandekerckhove admits that it is plausible that children before the critical age of four are capable, through merely procedural and semantic memory, of encoding personally-experienced events, though they are not *remembered* as such. This is shown by the imitative behaviour that develops early on. Infants will effectively reproduce previously experienced events through imitation, but do not retrieve the events themselves. Vandekerckhove suggests that infants form scripts (i.e., patterns of behaviour) for routine activities; they appear to 'remember' an earlier event when they perform those scripts, though she admits that much more empirical work needs to be done to verify this hypothesis (p. 11).

[16] But see Peterson (2002) on the idea that episodic memory is a gradual development beginning as early as the second year. I review this article below.

To test this account, Picard, Reffuveille, Eustache, and Piolino (2009) compared recent memories of school-age children with remote memories encoded during the period associated with infantile amnesia (in the first five years of life). They used the remember/know paradigm to test episodic recall,[17] and also measured frequency of rehearsal, emotion, and visual mental imagery. These measures were used in light of Conway (2005), who reported that in adults, emotions and visual mental imagery are most closely associated with EAM, and that rehearsal of memories can reinforce phenomenal details.

The authors found that recall of remote memories did not vary much with age, but that memories of events encoded in the first five years were less numerous, less episodic, and associated with less imagery than those encoded later. None of the participants could recall remote memories from the first two years. These results led Picard *et al.* to conclude, as did Vandekerckhove, that children *know*, but don't *remember*, events from the first five years of life. The existence of compelling competing accounts, and from my own observations of children during some limited piloting, I am sceptical of Picard *et al*. and Vandekerchove's claim. The results do not suggest a complete absence of episodic memory, though they do suggest that episodic memories are fewer and less detailed, which is consistent with the alternative view which I will now describe.

An alternative view on infantile amnesia holds that young children do have episodic memories, and the ability to access memories from as early as one to two years is reduced but not entirely eliminated after infantile amnesia sets in (Peterson, 2002). A number of studies demonstrate that adults can retrieve memories from before the age of two (Eacott & Crawley,

[17] See the introduction of this section for an overview of the remember/know paradigm. A more detailed description is also given in the Design and Procedure section in the *Study Proposals* chapter, as I make use of this task in my slub study.

1999; Howes, Siegel, & Brown, 1993; MacDonald, Uesiliana, & Hayne, 2000; Rubin & Schulkind, 1997; Usher & Neisser, 1993; Weigle & Bauer, 2000). In some of these studies, reports were compared with parents' reports, which mostly corroborated participants' reports.

Peterson acknowledges that such early memories have a tendency to be of short periods, infrequent, and of relatively poor detail. She does not, however, deny the episodic nature of these memories, as Vandekerckhove does. Peterson argues that infantile amnesia should not be considered an all-or-nothing phenomenon, as it is usually presented in the literature (see Pressley & Schneider, 1997; Rubin, 2000). Peterson cites several studies (e.g., Bauer, Wenner, Dropik, & Wewerka, 2000; Howe & Courage, 1997; Schneider & Bjorklund, 1988) claiming that the development of memory systems between infancy and preschool years is marked by more continuity than discontinuity. She notes that the concept of infantile amnesia was originally derived from adult studies of childhood memories, the implication being that the discontinuity only appears when early childhood is compared to a much later life period. There is no sudden change in the transition from infancy to preschool where the child begins to remember episodically.

The disparity in reports of onset of episodic memory is important to note, because it indicates a problem with the methods used to measure episodic memory. Since I want to examine the relationship between episodic memory and self-awareness, it is vital that I use a reliable measure of episodic memory. Episodic memory is difficult to pin down, however, because many supposed indicators of episodic memory lead to ambiguous interpretations.

As Bauer *et al.* (2000) notes, measures of episodic memory usually consist of verbal tests (e.g., "Where/when was the birthday party?") which can be misleading. For instance, when 5-year-old children are asked to act out an event rather than describe it verbally, they provide much

more information. Non-verbal measures are also problematic, however, due to ambiguity in the interpretation of non-verbal behaviour. To get around this difficulty, Bauer *et al.* note that, "what is needed is a nonverbal analogue to verbal recall: The mnemonic behaviour must be derived from a task that engages the same cognitive processes as those involved in verbal recall, yet does not require a verbal response." (p. 5) Based on previous research (Bauer & Wewerka, 1995, 1997; Mandler, 1990; Mandler, & McDonough, 1995; McDonough, Mandler, McKee, & Squire, 1995), Bauer *et al.* decided to use a paradigm known as *elicited imitation*. In this paradigm, the experimenter uses props to depict a sequence of events. The participant is then invited—after an optional delay—to imitate the events with the props. Bauer *et al*. added a 'baseline period' where participants were allowed to play with the props prior to the action depiction. This was to allow the experimenters to record any propensity to perform the target actions, to ensure that the participant is actually imitating the experimenter and not performing the same actions by coincidence. Bauer *et al.* found that infants as young as six months could successfully reproduce three-step sequences of events, although this ability decreased greatly when a 24-hour delay was imposed between demonstration and imitation.

Bauer *et al*.'s method seems promising as a measure of episodic memory, but it suffers from its own ambiguities. Nelson and Fivush (2004) correctly point out that elicited imitation studies do not conclusively demonstrate that children *re-experience* the staged events. The fact that children repeat the task in the right temporal sequence might seem to suggest episodic memory at work, especially in light of the fact that cases like K.C. are unable to recall temporal details of events. As I will indicate later in the review, however, children can rely on mechanisms other than episodic memory to encode temporal sequences. For example, they may encode *how* to perform the task in procedural memory, such that they do not need to re-

experience the event to replicate it. Temporal sequence may be necessary to demonstrate episodic memory ability, but it is not sufficient, so a less-ambiguous test is needed. Admittedly, the quality of *re-experiencing* associated with episodic memory is difficult to assess due to its highly subjective nature.

Relationship between Episodic Memory and Self-Reference

I now turn to a selection of studies that suggest a link between episodic memory and self-reference. These studies are of particular relevance for my first study, described in the next section. They do not specifically address the relationship between episodic memory and SAS consciousness—that is, they do not focus specifically on *non-ascriptive* self-reference—but they provide some guidance for my first study. Although my own studies only make use of behavioural measures, here I will examine both behavioural and neurocognitive research; the results of neurocognitive studies can be used to support the theoretical claim that episodic memory and self-reference are linked, although they don't tell us much about the nature of that relationship.

Craik *et al*. (1999) provide some neurological evidence of the relationship between episodic memory and self-reference. Previous studies (see Craik *et al.* for a review) demonstrate that episodic memory retrieval is primarily associated with the right prefrontal cortex, the same general area thought to be associated with representation of self. The authors wanted to test the hypothesis that the association of episodic memory retrieval with the right prefrontal cortex *is due to* representation of self.

It has been shown that people remember words better when they are used in a question about themselves than in generic questions (Craik *et al.*, 1999). Someone is more likely to recall the term “arrogant,” for instance, when asked earlier if it describes them than when they are

asked if "arrogant" means the same as "pompous." In the latter case, no explicit reference is made to self. In a PET (positron emission tomography) study, participants were given a word-recall task using self-referential, other-referential, or generic questions about adjectives. Activation of the brain during retrieval of non-self-referential adjectives was compared with activation during retrieval of self-referential adjectives. Analysis of the PET data showed that activation during the self-referential condition occurred in areas thought to be associated with episodic memory retrieval. Based on this result, the authors concluded that the self-concept (i.e., SAO) is a necessary component of episodic memory. As with any PET study, however, the localization of these activations is coarse-grained, so there is reason for caution in accepting the authors' claim. Neuroscience literature provides some support for the claim that episodic memory and self-awareness occur in roughly the same brain region, but this alone does not indicate that self-awareness is required for episodic memory.

A number of behavioural studies do provide such indicators. One such indicator is the "self-reference effect" or SRE, a behavioural measure of the effect that Craik *et al*. observed. Operationally, SRE is the tendency to recall items better when those items are associated with self at the time of encoding (Symons & Johnson, 1997). Typically, the tests involve semantic recall, so the study does not specifically address the relationship between episodic memory and self-awareness. I claim that SREs reflect an improvement in *episodic* recall, and that semantic recall improves due to the fact that many episodic memories have semantic components. If my claim is incorrect, then we should see no improvement in episodic recall corresponding to the improvement in semantic recall.

Initial scepticism about the existence of a distinct self-reference effect focused on a confounding factor—namely, that all of the self-referential items involved a person while all the

semantic ones did not. It was thought that the effect might be due to an association with people in general, not specifically self (Symons & Johnson, 1997). While some studies (e.g., Bower & Gilligan, 1979; Kuiper, 1982; Kuiper & Rogers, 1979) failed to find SREs after controlling for person-reference, many more did, as revealed by Symons and Johnsons' meta-analysis of studies on SRE. The authors attribute the seemingly conflicting findings to a failure to distinguish between familiarity and intimacy with the person referenced. They note that there is considerable overlap between self-reference and reference to an intimately-known other, since the other is highly self-relevant, so it would not be surprising if reference to an intimately-known other generated an SRE. Merely *familiar* others—such as celebrities—are not self-relevant, so we would not expect reference to them to produce SREs. After eliminating those studies that failed to make the familiar/intimate distinction, Symons and Johnson (1997) found that only intimacy predicted the magnitude of SREs.

Symons and Johnson's study limited the analysis to those studies that used a specific experimental paradigm. The paradigm consisted of posing a question to the participant, then presenting a word (the stimulus). The experimenter would either ask if a particular trait described *x*, where *x* was either the participant (self-reference condition) or someone else (other-reference condition), or ask whether or not the trait adjective meant the same as *y*, where *y* was some other adjective, either synonymous with the trait adjective or not ('semantic encoding' condition). The authors also limited the analysis to only those comparing self-reference manipulations to semantic manipulations, or self-reference to other-reference manipulations.

Symons and Johnson found that across studies (with some inconsistencies), participants recalled better in the self-reference condition than in the other-reference or semantic encoding conditions. It should be noted that for the purposes of my studies, what is important is not that

self-reference has a greater effect on memory than other strategies, but that it is a distinct effect. As will become apparent later, the possibility of an SRE makes the idea of self-attention—and corresponding effects of *distraction* from self—more plausible.

I hope to find a link between self-attention and episodic memory, as such a link would support my claim that SAS consciousness is required for the formation of episodic memories. Symons and Johnson's study suggests that this is a reasonable hypothesis, as it at least indicates a link between memory in general and self-reference. As I mentioned earlier, I explore the specific relationship between the SRE and episodic recall in my first study.

Effects of Inattention on Episodic Memory and SAS Consciousness

To my knowledge, there has been no empirical research done to establish when SAS consciousness is missing. This is unsurprising, as very little research has been done on SAS consciousness in general. Since I claim that the absence or relative lack of SAS consciousness results in a corresponding decrease in number and detail of episodic memories, a good place to start my investigation is the literature on conditions under which episodic memory is hindered.

The bulk of research on the effects of attention on episodic memory has been done by Gardiner and colleagues. Gardiner (2001) shows that episodic remembering depends on the cognitive resources available at the time of encoding. Participants who completed a divided-attention task demonstrated lower ability at recalling target items. Gardiner *et al.* (2006) discovered that under divided attention, semantic memory could take over the role of episodic memory. However, both semantic and episodic memory improve when slow, effortful processing is applied at the time of encoding, challenging the widely-held view that knowing is automatic and relatively effortless, while remembering is a more conscious and effortful process.

As mentioned earlier, I claim that episodic memory is dependent upon SAS consciousness—being necessarily self-referential—while semantic memory is not. If my claim is correct, then I expect that if one's attention is directed away from SAS during a given event, then it will be harder to recall episodic memories of the event. Likewise, encoding of non-self-referential semantic memories should not be affected during lapses of SAS consciousness, all else being equal. To test this claim, I attempted to induce lapses in SAS consciousness, and then measured episodic and semantic recall a short time later. Given the nature of the methods I describe in my studies, I expected that rather than lapses of SAS consciousness persisting for long, continuous periods of time, they would last only for brief moments interspersed throughout the trials. As such, I expected attenuation rather than complete absence of episodic memories; specifically, I expected a reduction in the recall of episodic details during distracting trials, relative to non-distracting trials.

There are many possible reasons why the encoding of episodic memory might be affected by divided attention. The phenomenon of 'inattentional blindness,'—where a person fails to be aware of some visual stimulus when her attention is fully occupied by other visual stimuli (Mack & Rock, 1998)—may be at work, as much of the detail we recall episodically is often of a visual-spatial nature. It is also possible that encoding of episodic memory is hampered because much of the visual stimuli relevant to the episode fail to be attended to due to distraction. As yet, no studies have been done to support either suggestion, and the first study that I describe in the next section is designed to determine what mechanism is responsible for the deficiency in episodic memory encoding.

Since there are no studies on the effects of lack of attention to SAS on episodic memory, I need to establish the plausibility of an effect by looking at parallels, that is, the effects of lack of

attention on other cognitive faculties. Much has been written, for instance, on the effects of attention on the ability to enumerate. For example, Vetter *et al.* (2008) found that under high attentional load, participants' ability to enumerate suffered significantly. In the study, participants were given two tasks. In the high-load condition, participants were asked to complete two different tasks at the same time. The primary task required participants to identify the colours in a small diamond-shaped area in the center of the screen, while the secondary task consisted of a set of Gabor patches,[18] varying in number, arranged in a circle around the center. Participants were asked to report the number of patches. When participants were asked to perform both tasks, the accuracy of their responses on the secondary task dropped considerably. This result indicates that the limit on attentional resources can easily be exceeded when the participant is asked to perform just two relatively simple tasks, at least in the case of single-mode attention. An important question is whether or not self-attention is affected by other sensory modes. If self-attention requires the same resources as, for example, visual-spatial or auditory attention, then loading those resources should result in a decrease of self-attention. There is some evidence of such resource sharing (Carver & Scheier, 1978; Davis & Brock, 1975; Ganellen & Carver, 1985), however these studies only explored *increase* of self-attention due to visual, auditory, and linguistic stimuli. I turn now to a review of these studies.

In order to study the effects of attention on self-awareness, we need to compare cases of little or no self-awareness with cases of normal or high self-awareness. There has been little or no research done on 'low self-awareness.' There is, however, some research on the elevation of self-awareness, broadly construed—that is, without separating SAS and SAO. The following studies are important for two reasons: they indicate one possible measure of self-awareness and

[18] A Gabor patch is a sinusoidal grating, often filtered in such a way that the edges blend in with the background. See Mathôt (2010) for examples generated by a freely available web-based tool.

demonstrate that attention to self can vary—although they fail to tell us whether or not self-attention can be entirely absent.

Davis and Brock (1975) showed that being recorded by a TV camera reliably increased self-attention. Prior to the introduction of the camera, participants completed fake creativity tests, and then were given positive or negative feedback, or no feedback concerning their results. They were then shown foreign-language sentences (in languages unfamiliar to them) and a list of English pronouns, and asked to select the English pronoun corresponding to an underlined foreign pronoun. Participants who were exposed to a TV camera selected first-person pronouns more often than those who weren't exposed, regardless of the amount or kind of feedback given about self. The authors interpret this result as indicative of increased self-attention. This study is interesting for the purposes of my experiments, because it makes use of the first-person pronoun as an indicator of increased self-attention. I am primarily interested in cases of decreased self-attention. If Davis and Brock's interpretation is accurate, I expect to see a corresponding decrease in use of the first-person pronoun under conditions of low self-attention.

Similarly, Carver and Scheier (1978) demonstrated that exposure to a mirror encourages people to focus attention on self-related thoughts, feelings and attitudes. In Carver and Scheier's study, self-focus was evaluated by having participants complete sentence stems. The content of completed sentences were then scored as reflecting focus on self, world, both (ambivalence), or none. Scoring was done using Exner's (1973) criteria, where a self-focused response was judged as one that was clearly related to self with little or no regard for the external world (including other people). To use an example of Carver and Scheier's, given the stem 'it is fun to daydream about:' a completion such as 'my success' or 'being loved' was scored as self-focused. Contrariwise, responses that focused on other people or objects in the world—that according to

Exner's (1973) criterion show 'concern for real things or people' (p. 440)—were coded as externally-focused, such as 'marrying Tom' or 'giving a party for friends' (Carver & Scheier, 1978, p. 326). Ambivalent responses were coded as both, while vague responses were scored as neutral. Carver and Scheier found that participants who completed sentence stems in front of a mirror wrote more self-focused completions than those who did not. It should be noted that the authors ran the same experiment using the presence of an audience instead of the presence of a mirror, and found a similar result.

In later work, Ganellen and Carver (1985) found a distinct effect on ability to encode memories incidentally—that is, to encode a memory of an ostensibly task-irrelevant piece of information—when participants are asked a question about a word involving themselves; more specifically, when participants were asked how strongly they felt that a certain word described themselves, they were far more likely to remember the word later. The researchers demonstrated that this effect was independent of the emotions generated by the word, as well as its perceived importance or distinctiveness. This result strongly indicates a distinct 'self-reference' effect.

These studies demonstrate that one can direct more or less attention to self—broadly construed as including both SAS and SAO—and Ganellen and Carver's study demonstrates that there is a distinct effect of self-reference over and above other plausible mechanisms for improving memory. Unfortunately, these studies fail to indicate whether or not it is possible for attention to be entirely directed away from self. The first study that I describe in the following section explores the idea that one can be distracted from self—at least to a large extent. My assumption is that if one's attention is focused on non-self-related contents, then less attention will be paid to SAS, and I should see a corresponding reduction in episodic memory recall. I attempted to distract participants from SAS by asking the participant to perform a challenging,

abstract task (a game of Tetris) while simultaneously performing an episodic memory encoding task (listening to a short auditory narrative).

Summary

In this section I have examined the few existing studies that concern the nature of SAS consciousness and its relationship to other cognitive faculties. Since there have not been any studies clearly focused on SAS consciousness, my review focused on work that provides evidence of the three features of SAS consciousness that I identified in the *Conceptual* section. The studies on mastery of the first-person pronoun do an adequate job at establishing the upper limits of the onset of SAS consciousness, but tell us little about the earliest time of onset. Due to the young age at which the first-person pronoun mastery occurs, and assuming that the onset of SAS consciousness precedes that event, it was impractical for me to investigate the question of the earliest period of onset. Instead I focused my efforts on determining the period of development when children begin to attribute instances of SAO to self, an ability that the results of Povinelli's study suggest is underdeveloped long after first-person pronoun mastery.

I also identified episodic memory as the clearest example of a faculty reliant upon the remaining features of SAS consciousness, universal availability and non-attributive reference. The studies on episodic memory that I have examined suggest that episodic memories depend on having SAS consciousness extending into the past. Just as SAS consciousness is required to experience the present as *happening to* you, so it is required for recalling previous events from the standpoint of the earlier experience. The relationship between episodic memory and SAS consciousness is relatively unexplored, but existing studies on declarative memory in general and self-reference can provide some guidance for future studies. Ganellen and Carver, for instance, demonstrated an interesting relationship between memory and self-reference in general, but did

not specifically address the relationship between *episodic* memory and self-reference. One of the goals of my studies—detailed in the following section—was to explore this more specific relationship.

Empirical Studies

Introduction

In the *Conceptual* section, I outlined a number of properties that SAS consciousness has. First, reference to SAS must be indexical. Second, the notion of SAS is part of the *structure* of representation, so it is universally available to us in experience. Last, SAS consciousness is non-attributive, meaning that in referring to SAS, I do so without needing to attribute anything to myself. These characteristics are the extent of what we know about SAS consciousness and the limit to what we can know about SAS consciousness through reflection and introspection. We want to know about the psychological structure of SAS consciousness, however, and that requires experimentation. In the previous section, I gave an overview of the current state of the art in the scientific study of consciousness of self. Here, I give a detailed description of two studies I designed to test or explore some of the claims that follow from the framework I outlined in the *Conceptual* section. The first study is designed to examine whether or not there are relationships among attention, self-awareness, and episodic memory, and if so, what the nature of these relationships are. The second study is a preliminary investigation into the development of SAS consciousness in children between the ages of 3 to 5 years. Due to task-related constraints and practical issues of recruitment—described later in this section—it was not possible to obtain a large enough sample to perform meaningful inferential analyses. The study did, however, demonstrate some interesting trends that would be worth investigating further.

For each study, I give a brief overview connecting existing literature to my studies, followed by a description of the methods I employed to answer my research questions. Finally, I present the results of my analyses, discuss my findings and make some suggestions for future research.

My overall goal is to discover details of the structure of SAS consciousness, and determine how SAS consciousness relates to other cognitive faculties. My project addresses the following research questions:

1) Does the encoding of episodic memories depend on SAS consciousness?
2) By what age are typically-developing children capable of recognizing themselves as SAS?
3) When does SAS consciousness become fully integrated with SAO consciousness?

To answer these questions I ran two studies, the first involving adult participants and the second involving children. The first study, addressed in the following section, addresses the first question, and represents the bulk of my empirical investigations. The second study is a preliminary investigation into the last two questions and comprises the final section. The hypotheses corresponding to these questions are outlined at the beginning of each section.

Study I: Effects of Attentional Load on SAS Consciousness

Hypotheses

My first study explored the effects of attentional load on SAS consciousness in adults. The study was designed to test Hypothesis 1: *The encoding of episodic memories is possible only when one is conscious of SAS.*

This hypothesis seeks to further our knowledge of the details of the structure of SAS consciousness. While we can be confident that the distinction between SAS and SAO exists, we

want to know more about what abilities specifically depend upon SAS consciousness but are independent of SAO. Only by learning more about these dependencies can we understand how SAS consciousness influences and is influenced by other faculties, and thereby develop a more detailed model of it. One faculty that appears to depend on SAS consciousness is episodic memory.[19] With Gardiner (2001), I define episodic memory in general as memory of events, as experienced from the perspective of the person, and recalled in the same temporal sequence as the original experience. I suspect that—since all episodic memories are recalled (or recollected) specifically as *having been experienced by* self—if SAS consciousness is absent the person will be unable to encode episodic memories. This quality of episodic remembering as recalling from the standpoint of the previous experiences, I call the *self-related* component of episodic memory.

This hypothesis assumes that one can be conscious without being conscious of SAS when one's attention is fully engaged. Originally, I intended to state this assumption as a separate hypothesis, but doing so created circularity, as it was precisely the reduction in episodic memory that would be indicative of consciousness without SAS. This assumption is contrary to the claim, widespread among philosophers, that to be conscious is also to be conscious of self, such that to be conscious of a mailbox, for example, is necessarily to be conscious of *being* conscious. My study assumes that under certain conditions, SAS consciousness can be reduced without a loss of consciousness in general. I postulated that an absence of self-reflection combined with adequate attentional load would result in reduced SAS.

Introduction

To test this hypothesis, I combined introspective and behavioural measures in the hopes of finding convergence among them. I reasoned that if SAS consciousness was a principal

[19] See the review in the previous section for more on how these are connected.

component of episodic memory, then encoding of episodic memories should be restricted when SAS consciousness was inhibited. In order to inhibit SAS consciousness, I loaded the participant's attention with a challenging, abstract task (a simplified game of Tetris).

According to attentional load theory (Lavie, 2004; 1995), there are a maximum number of items that a person can attend to at any given time. When attention is maximally loaded, the person fails to be conscious of additional stimuli presented in the same mode. For example, if the person fully loads her visual-spatial attention by counting the number of times a basketball is passed between players, she may fail to attend to fairly obvious visual-spatial changes, such as a gorilla walking through the middle of the scene. This deficit is called "inattentional blindness" (Mack & Rock, 1998).[20] Invariably, this term is limited to the loading of visual-spatial stimuli, but it is reasonable to suppose that a similar effect occurs within other modes, and there is some evidence that distraction works across modes (Driver & Spence, 2004; Sinnett, Costa, & Soto-Franco, 2006).

I postulated that the act of self-reference would take up attentional resources. If these resources were shared with other faculties—as the self-reference effect suggested (see discussion of this effect in *Research on Indicators of SAS Consciousness,* in particular the overview of Symons and Johnson, 1997)—then fully loading those resources should have resulted in reduced SAS consciousness. I expected that the loss or reduction of SAS consciousness would in turn result in a reduced ability to encode episodic memory during the time of encoding, based on hypothesis (1), that SAS consciousness was required for the encoding of episodic memory.

I have argued that SAS is not an object of experience, but part of the structure of representation, so it may seem odd to suggest that it would be affected by attentional load in the

[20] See the preceding review for more on inattentional blindness and attentional load.

same manner as the contents of experience. If SAS is not part of the content of experience, why should I expect attentional load to affect it? My claim is that while SAS is part of the structure of representation, an additional cognitive mechanism is needed to explicitly represent SAS.

To test my hypotheses, I fully loaded participants' attention using a well-known video game (Tetris) with which I expected them to be familiar, and manipulated the degree of self-reflection required between two groups—low self-reflection (LSR) and high self-reflection (HSR). In the LSR group, I played a narrative that featured a familiar protagonist, while in the HSR group I played a narrative where the protagonist's name was replaced by the second-person pronoun (*you*), to force the participant to reflect upon herself (see *Materials and procedure* below for a more detailed description of the task). There were no other differences between the LSR and HSR groups, as I wanted to introduce as little variation as possible. I did not expect to entirely eliminate self-reflection in the LSR group, but by comparing that group with an HSR group where there was a very high probability for self-reflection, I expected to see some effect if episodic memory was affected by changes in SAS consciousness. If episodic memory was not affected by SAS consciousness, or SAS consciousness did not share resources with sensory modes, I would not expect an effect.

Since it is known that attentional load restricts the formation of episodic memories due to factors unrelated to the dependent variable (see Gardiner, 2001; Gardiner *et al.*, 2006), it was important to give both groups an equal degree of attentional load, and load the same resources.

Method

Participants.

In total, 64 participants (37 female) between the ages of 17 and 50 (only one under the age of 18) were recruited from the undergraduate psychology pool at Carleton University, or

from my own circle of acquaintances and colleagues at Carleton University or in the Ottawa area (the latter were recruited for only the pilot phase of the study). Informed consent was obtained from all participants. SES data were not recorded or controlled for. One participant was excluded from analysis after revealing a history of concussions that greatly affected her ability to remember. One was excluded after her data were accidentally recorded over by data from another participant tested the same day. One was excluded because she skipped a question on the Cognitive Failures Questionnaire (described below), which I failed to notice during the session. One participant (the first in the full study) was excluded after revealing that she had been looking at the keyboard to help jog her memory (and in all subsequent trials I ensured that the laptop was closed when not in use).

Ten participants were assigned to the pilot phase of the study (described in *Methods – Pilot study*, below) and 10 were assigned to a special follow-up study (described in *Methods – Follow-up study*, below) conducted midway through the full study. This follow-up was intended to ensure that the central task (the Tetris/Auditory Temporal Sequence task, described below) was not too easy. Participants in all three phases were randomly assigned to either the control group (high self-reflection or HSR group, explained below) or test group (low self-reflection or LSR group, explained below). An equal number of participants in the pilot study and follow-up study were assigned to each group. In the full study, 19 participants were assigned to the HSR group and 21 were assigned to the LSR group.

Forty participants (23 female) between the ages of 18 and 39 (M = 21 years, SD = 4.1 years) were included in the final analysis, all recruited from the undergraduate psychology pool at Carleton University.

Design.

All participants were tested individually by a single experimenter (the author). Participants completed six tasks in one session, lasting approximately 45 minutes to an hour. The tasks consisted of the Tetris/Auditory Temporal Sequence (T/ATS) task, the Visual Attentiveness/Sustained Inattentional Blindness (VA/SIB) task, the Self-Referential Adjectives (SRA) task, the Free Recall task, the Remember/Know task, and the Cognitive Failures Questionnaire (CFQ). See *Materials and procedure,* below, for a detailed description of each of these tasks.

Task order was randomized to control for order effects, though one exception was made for practical reasons. Specifically, the VA/SIB was always placed between the encoding and recall phases of the SRA, to provide a constant interval across participants, and to keep the participant occupied. I treated the combined SRA and VA/SIB task as a unit and used a pseudo-random number generator (a custom Python script) to determine the order of tasks.

Materials and procedure.

I presented all visual stimuli to participants on a laptop computer screen, and auditory stimuli through the laptop speakers. The same laptop was used for all participants. A neutral blue background was used for all visual stimuli, as this colour reduced the reflectivity of the screen. This was an important detail, as I wanted to control opportunities for self-reflective thought, and physical self-reflections were known to encourage self-related thought (Carver &Scheier, 1978).

Self-referential adjectives/Abstract visual distraction.

In this dual-task paradigm, I tested for the self-reference effect (SREs) on recall of word lists while the participant was distracted by an abstract visual task. Numerous studies have demonstrated this effect using the “self-referential adjectives” paradigm. In this task, the

participant is presented with a set of words, and asked either how well the word describes the participant, or how well it describes a well-known public figure (Canadian Prime Minister Stephen Harper). All words were adjectives generated by the online Word Generator random word generator (2011; available at http://www.wordgenerator.net/random-word-generator.php). Adjectives were added to the word list as they appeared, with no other selection criterion applied. I used the same list of words, presented in random order, for all participants. All linguistic stimuli were presented on a computer screen as white text against a black background. Questions were presented in auditory form as recordings of the experimenter's voice. To ensure that participants would not pick up on subtle auditory differences between questions, I recorded the leading phrase "how well does this word describe…" separately from the words "Stephen Harper" and "you," then inserted the trailing words digitally.

At the same time, the participant is asked to control a ball on a computer screen, and try to keep the ball (a small white filled circle) between two lines on the screen using arrow keys.[21] A sequence of short white vertical lines extending to the left and right of the screen represented the distance from the center of the screen. This distractor task remained on screen during all trials, without interruption.

The task consisted of 60 trials; in each trial, the program presented the participant with a novel word, presented in random order. No words were repeated. At the beginning of each trial, a white noise mask appeared over the space where the words displayed, for 500 ms. During each trial, an audio recording played, asking either "how well does this word describe you?" (self-referential question) or "how well does this word describe Stephen Harper?" (other-referential question). The question was chosen semi-randomly by the computer program; the program was

[21] Not to be confused with the Tetris game used as a distraction in the central task, described later.

designed such that exactly 30 of each question would be asked. Question order was randomly determined, but if one question repeated 30 times, the remainder of the trials would use the other question. This was to ensure that the participant would hear the same number of self- and other-referential words. I instructed the participant to use the number keys on the keyboard to specify how well they thought the word described a particular person (revealed during the trials), on a scale of one to four, where one was "very poorly" and four was "very well". The program allowed a maximum of 10 seconds per question. If the participant answered the question, or 10 seconds elapsed, the program would automatically move on to the next trial.

To distract the participants while performing the description task, they were asked to keep the white circle at the horizontal centre of the screen using the left and right arrow keys. Three distance markers consisting of short white horizontal lines extended from the center to either side of the screen, and the circle would begin moving to either the left or right (randomly determined) at the beginning of the task. The participant could move the circle left or right, but the speed of each movement was randomly determined such that the circle would continually drift away if the participant did not attempt to correct it. Participants were told that they would be scored based on how close the circle stayed toward the center of the screen, however no score was displayed or recorded. The task was intended solely as a distraction.

After a three-minute intervening task (the VA/SIB, described below), I provided the participant with a sheet of paper and asked her to write down, within a five-minute period, as many words from the list as she could recall. The participant was scored on the number of self-descriptive words she could recall, divided by the number of other-descriptive words she could recall. A score of '1' indicated that the participant recalled self- and other-descriptive words equally, a number approaching '0' indicated that she recalled more other-descriptive words than

self-descriptive ones, and a number greater than '1' indicated that she remembered more self-descriptive words (the highest possible score would be 30). The participant was also scored on the total number of words she recalled (out of 60, converted into a percentage), as well as accuracy of recall (number of correctly-recalled words divided by total number of words listed).

Visual attentiveness/Sustained inattentional blindness.

This task is a standard sustained inattentional blindness paradigm (Most, Simons, Scholl & Chabris, 2000); it is a 'sustained' inattentional blindness task, meaning that it measures inattentional blindness to an unexpected, but prolonged, presentation of a stimulus. Most *et al*. (2000) used the paradigm to study the effects of spatial proximity on inattentional blindness; they found the task to be highly effective at inducing such blindness.[22] I used it simply to control for individual variance in susceptibility to inattentional blindness. Some participants were likely to be better at attending to multiple stimuli than others, and I wanted to ensure that any differences I observed between LSR and HSR conditions were due to the effect of self-attention and not simply a reflection of general ability to attend to multiple targets at once.

On a computer screen, I displayed a series of moving letters—capital 'L's and 'T's—that moved in random directions and occasionally crossed the center of the screen. A fixation cross remained in the center of the screen during each trial, as did a thin (two-pixel wide) horizontal blue line running through the fixation cross. I asked participants to mentally keep track of the number of times the 'L's touched or crossed the center line, but to ignore the 'T's. In some trials, an unexpected event would occur midway through. In the unexpected event, a grey cross moved horizontally across the center of the screen at a fixed speed, from center right to the center left,

[22] Specifically, Most *et al*. (2000) found that spatial proximity did have an effect on inattentional blindness, such that participants were more likely to detect targets that moved closer to the line attended to than others (details of the task are explained later in this section). Even when the target moved along the line, however, the authors reported a high incidence of inattentional blindness—only 47% of participants detected the target.

passing through the fixation point, travelling across in five seconds. I presented one practice trial and five actual trials, each 15 seconds long. The practice trial and first two actual trials did not include the unexpected event, but the remainder did. At the end of each trial, participants were prompted to enter the number of times the 'L's crossed the center line. At the end of the remaining trials, participants were also prompted on whether or not they saw anything unexpected, and if so, asked to describe what they saw.

Participants were scored on how many of the final three trials they report seeing the cross, receiving one point in each trial for claiming to notice something unexpected, and one point in each trial for correctly identifying the cross, for a maximum of six points. Very high (5 or more) or very low scores (1 or fewer) on this task were taken into consideration when analyzing performance on the episodic free recall task, described below.

Tetris/Auditory temporal sequence.

This was a dual-task paradigm used in combination with the episodic memory recall task to test the hypothesis that a reduction in SAS consciousness (in the form of self-related thoughts) leads to a reduction in episodic memory formation. The aim of these tasks was to produce circumstances under which the participant would be less able to form episodic memories of events. In the episodic free recall task, described in a later section, the participant was tested on her ability to recall details of the presented events.

I distracted the participant from self-related thoughts using a game of Tetris while presenting a narrative depicting a sequence of events during a trip to a cabin. The game of Tetris was chosen because it was expected that most people would be familiar with the game, and comfortable enough with the game to easily become immersed in the task. Even if participants were not familiar with the game, the controls (which I explained to them and trained them on in a

practice trial) were fairly straightforward and the mechanics of the game very simple (yet challenging).

Before the task began, I asked the participant if she was familiar with the game of Tetris. In both conditions, I told the participants that during the game, an auditory narrative would play over the speakers. I further told them (falsely) that the narrative was intended to distract her from the Tetris game. This slight deception was necessary to ensure that the participant focused on the Tetris task, since if she knew that the real purpose of the exercise was to study her recall of the narrative, she might deliberately devote less attention to the game to concentrate more on the narrative. To further encourage the participant to focus on the game, I told her to try to achieve as high a score as possible, and told her that losing the game would cause the game (and their score) to reset. Even if the participant was familiar with the Tetris game, I administered a practice/training trial consisting of a 30-second game of Tetris, with no narrative in the background. This was to ensure that the participant was comfortable with the controls, and would not be struggling to learn the basic gameplay during the testing trial.

During the testing trial, both conditions began with a white 'READY' text displayed on a blank black screen for 500 ms. The Tetris game then appeared on the screen and the game began. Thirty seconds later, the narrative started to play over the speakers. In the control condition (HSR), the narrative featured the pronoun "you" in place of the protagonist's name (since the narrative was a recording, it was not possible to use the participant's name). In the test condition (LSR), the narrative featured a well-known but not personally-known protagonist (Stephen Harper). A transcript of the narrative, presented in first-person form, is included in Appendix F. The game ran for exactly five minutes, regardless of the participant's performance, and there was no limit on the number of times it could reset during that period.

This task relied upon the assumption that the Tetris game distracted the participant from self-reflection. It is important to note that the task did not need to preclude *all* self-reflection, so long as it *reduced* self-reflection. I expected an average lowering of SAS consciousness in the LSR condition; I fully expected that participants would intermittently engage in self-reflection, but that this would occur less frequently in the LSR condition.

After running several participants, I grew concerned that the task might be too challenging, so I ran 10 additional participants (not included in the number of participants listed above, or in the final analysis) without presenting the Tetris game as a distraction. Participants were simply asked to listen to the narrative, but were not told why. As usual, half of the participants were assigned to the HSR condition and half to the LSR condition. During these sessions, no other changes were made to the materials or procedure.

Earlier studies demonstrated that novelty increased the likelihood of episodic encoding (Tulving & Kroll, 1995; Tulving, Markowitsch, Craik, Habib, & Houle, 1996); I wanted to ensure that LSR and HSR conditions were equally memorable in every respect other than self-relatedness. I expected a possible confound from the von Restorff effect, which is the effect of distinctive stimuli causing an item to be more memorable (see Cohen and Carr, 1975, for an example of this effect with respect to human faces). I wanted to ensure that the effect being measured was due specifically to *self*-reference and not merely to a sense of novelty. Therefore, I made stimuli in both conditions distinctive, with the only difference being that the HSR narrative made reference to the participant. The non-self-reflective narrative featured Canadian Prime Minister Stephen Harper, in keeping with Symons and Johnson's (1997) observation that descriptions of merely familiar people did not produce SREs; I wanted to ensure that, while distinctive, the stimuli would not trigger an SRE. Since it was unlikely that any of the

participants had a personal relationship with Stephen Harper, but were likely to be familiar with him, he was an ideal candidate.

Tetris 'dummy'.

Following the T/ATS (Tetris/Auditory Temporal Sequence) I told participants that they could take a five minute break, to allow the participant time to forget details from the narrative. Participants were instructed not to discuss the study with anyone during this break. In the process of testing, I discovered that many participants did not wish to take the break, and several were noticeably uncomfortable sitting in silence for five minutes, or would wonder aloud about the reason for the pause. As a result, I had to explain to participants that the break was a necessary part of the study, without revealing why. Out of concern that this might tip them off about the true purpose of the break, I decided to fill the time by asking them to play Tetris again, this time without anything happening in the background. This had the added benefit of adding consistency to the experience across participants. I also thought that if the participants were occupied with the game, they would be less likely to rehearse the details of the story. Data from this version of the game were not used in the analysis, since the task was introduced mid-study after already collecting data from 17 participants.

Episodic free recall.

Five minutes after the T/ATS, I asked participants to complete an electronic questionnaire designed to test their ability to freely recall details of the events depicted in that task. Free recall is thought to rely heavily on episodic traces more so than cued recall (Tulving, 1985; Perner, Kloo, & Gornik, 2007). Participants were not informed ahead of time that they would be asked to recall the narrative, to ensure that no cues could be given unintentionally. Since participants were

not informed about the task ahead of time, they were asked to sign an informed consent for the release of data at the end of the session.

The computer program presented 16 questions in text format, one at a time, presented in random order (see Appendix G for a list of questions asked). Each question was presented as black text against a neutral blue background. Participants could type in their answers and review them, but were informed that once they hit "enter" they would not be able to go back and change their answers. The order of the questions and the responses to the questions were recorded electronically, however the scoring of the responses relied to some extent on human judgment.

I scored participants on the number of details recalled correctly, as well as the total number of distinct details that they recalled about the experience, and the number of questions answered. For the purposes of measurement, a point was given for each word or phrase that provided a unique piece of information. Repeated mention of details or mention of a detail that was provided in any previous question or answer did not count toward the number of points given. A detail was deemed correct only if it matched very closely with the information in the story. When in doubt, I considered the detail incorrect. Since I was not blind to the hypotheses being tested, I employed a second rater to evaluate the reliability of the measure.

Remember/know.

This task is a well-established measure of episodic memory in adults, described by Tulving (1985; 1993). It was used to assess whether participants characterize their memories in terms of remembering or in terms of knowing. The paradigm is based on the idea that the conventional use of the term "remember" refers to episodic recall, whereas information recalled

through semantic memory alone is commonly thought of as "knowledge."[23] In this task, confidence scores are used as proxies for remembering and knowing. High confidence on a positive answer that a word was in a list suggests that the participant is actively remembering the task, while a moderate degree of confidence suggests that the participant knows it was in the list, but can't "remember" it episodically. No confidence in the answer suggests that the participant is guessing (i.e., not making use of either memory system to determine the answer).

I began by telling participants that I was going to show them a list of nouns, presented one at a time, and that they would have to make a judgment about whether the nouns were 'living' or 'non-living'[24] by pressing the blue or yellow button on the keyboard, respectively (coloured stickers were placed on the keyboard to indicate what keys to press). Following a very similar procedure outlined in Tulving and Kroll (1995), I then presented participants with two sets of 20 common dual-syllable, singular English nouns to study. All sets were drawn from a list derived from the Word Generator random noun generator (2011; available at http://www.wordgenerator.net/noun-generator.php). Nouns were added to the list as they appeared if they met the above criteria of being dual-syllable, singular, and common – otherwise there was no selection. These sets were drawn from a pool of four sets (a total of 80 words). I presented the words on the laptop display one at a time in random order, and presented each set four times, for a total of 160 words (160 trials). Each trial ended when the participant pressed either the blue or yellow button on the keyboard. Tulving and Kroll used twice as many items,

[23] See the preceding review in *Research on Indicators of SAS Consciousness* for more details on the theory behind the remember/know paradigm.

[24] These categories were open to interpretation. I wasn't interested in the judgments of the words themselves, only the ability to remember the words later. The fact that some words were ambiguous ('werewolf' for example, or things that were formerly alive, like 'cigar') may have helped keep participants engaged in the task.

but found similar results when they halved the number of items, so I opted to use a reduced number in the interests of shortening the duration of the task.

Following a short break (30 seconds), I presented both sets of nouns once (40 words total). For each noun, I asked participants to make a judgment about whether or not the word appeared in the previous list, this time by pressing either the green button or red button, respectively. I then presented the same set of 40 words again, and asked the participant to make living/non-living judgments using the blue and yellow buttons.

Two minutes later, I told the participant that I would now show her a 'critical study list'. I presented a second list of 40 randomly-ordered words (the 'critical study' list) containing a mix of one familiar set (20 words from the original list) and one unseen set (20 new words). The new words had no semantic relationship to the old ones (see Appendix I). I told participants that for each word, they were to indicate (via the keyboard) whether or not the word was in the original list. This part of the task took about two minutes.

Two minutes later, I presented all 80 words from the pool, one at a time. For each word, I asked participants to judge whether or not the word appeared in the 'critical study' list, and to rate the confidence of that judgment on a three-point scale—3 for very confident, 2 for fairly confident, and 1 for not at all confident. All answers were collected electronically via keyboard (*y/n* for the first question and *1/2/3* for the second). This part of the task took about five minutes to complete.

I calculated the mean confidence scores for all novel words ($confidence_{novel}$) and for all familiar words ($confidence_{familiar}$), and calculated an accuracy score by determining the percentage of correct answers on the 'yes/no' question, and generating three scores for each participant ($confidence_{novel}$, $confidence_{familiar}$, and accuracy). High confidence scores were taken

as an indication of episodic memory and a measure of recall ability. Moderate confidence scores were taken as an indication of semantic knowledge, and low confidence scores suggested that participants were guessing.

Cognitive failures questionnaire.

As the propensity for distraction could have accounted for differences in performance on the Free Recall task, the CFQ (Broadbent, Cooper, Fitzgerald, & Parkes, 1982. See Appendix H) allowed me to control for non-task-related distractions. A large number of studies (e.g., Kass, Beede, & Vodanovich, 2010; Cheyne, Carriere, & Smilek, 2006; Vom Hofe, Mainemarre, & Vannier, 1998; Larson, Alderton, Neideffer, & Underhill, 1997; Larson & Merritt, 1991; Tipper & Baylis, 1987) have shown that this questionnaire is a reliable indicator of distractibility. Together with the measure of inattentional blindness (the VA/SIB) I expected to effectively control for variability due to deficits of attention.

I asked participants to fill out the printed questionnaire, telling them that it was a questionnaire about "mistakes that everybody makes from time to time," which was also stated at the beginning of the questionnaire (see Appendix H). The title of the questionnaire was not revealed to the participants out of concern that they might have a negative reaction to the term "failure," or that this might prime them toward answering untruthfully.

The questionnaire consisted of 25 questions on a five-point Likert-type scale from zero to four. Participants could score a maximum of 100 points on the questionnaire. Questions identified the frequency of various cognitive errors—such as memory or perceptual errors—over the last six months. The questionnaire was almost identical to the original by Broadbent *et al.* other than a few changes to the language to reflect a Canadian audience (the original was written

for a UK audience). Participants were given as much time as they needed to complete the questionnaire, but this did not typically exceed a few minutes.

Method - Pilot study

Prior to testing the 40 participants above, I tested the design of the stimuli and tasks in a pilot study. As a result of the pilot, I decided to replace the central task (BC/TS, described below) with the T/ATS task described above. The pilot data were not used in the analysis, as substantial changes were made to the study, and no reliable analysis could be done as the design of the study was continually updated based on feedback from participants. The primary goal of the pilot was to troubleshoot the design of stimuli and tasks, however it became clear that the BC/TS was too difficult, leading to its replacement by the T/ATS.

Participants.

Ten participants (six female and four male) between the ages of 24 and 50 (M = 33 years, SD = 9.6 years, 1 missing) were recruited from my colleagues at Carleton University and circle of acquaintances. Participants did not receive compensation for participation but were provided with free parking when applicable. Informed consent was obtained from all participants. All participants spoke fluent English; of these, five were monolingual, one spoke fluent French, one knew some French but was not bilingual, and the remaining two knew some French plus one other language (German and Spanish) but were not fluent in those languages. I assigned five participants to the control condition (high self-reflection) and five to the test condition (low self-reflection). Since the purpose of this pilot study was largely to test stimuli, and substantial changes made to the stimuli prior to the full study, none of the data collected were used in the final analysis.

Design.

All participants were tested individually by a single experimenter (the author). Participants completed six tasks in one session, lasting approximately 45 minutes to an hour. All of the tasks listed above, with the exception of the T/ATS, Episodic Recall, and the Self-referential Adjectives/Abstract Visual Distraction task, were featured as described in the pilot. Task order was randomized using the same process described in the full study. The Backward counting/temporal sequence task described below was presented to the participant on a laptop computer using a custom Python script I wrote specifically for the study.

Materials and procedure.

Backward counting/Temporal sequence task.

In both conditions, the participant was asked to perform an abstract task that demanded her full attention. I used the paradigm described in Karlsen, Allen, Baddeley, and Hitch (2010), which was a relatively simple yet attentionally-demanding backward counting task. Backward counting is typically used as a working memory task, as it requires holding a particular item in mind while manipulating it—in this case, subtracting three. Working memory is the ability to keep items in short-term memory via rehearsal (Baddeley & Hitch, 1974).

Earlier studies (see Karlsen *et al.* 2010, for a review) demonstrated that this task places a heavy load on the attentional resources required to perform visual-spatial encoding tasks. To load participants, I asked them to verbally count backward by threes (e.g., 'two-hundred and forty-one, two-hundred and thirty-eight'), starting from a randomly-generated three-digit number announced at the beginning of each trial. I was not interested in the accuracy of participants' responses, as the task was intended simply as a distraction; however I did record accuracy in case I found an interesting correlation with another measure.

In both LSR and HSR conditions, I told the participant that for each trial, a three-digit number would be played through the speakers, and that as soon as she heard the number, she must verbally count backward by threes as quickly as possible until the word 'STOP' appeared on the screen. I also told her to repeat the process when she heard the next three-digit number (signalling the start of the next trial). I informed the participant that during each trial, a series of three pictures depicting a common event would be displayed one at a time, in temporal order, and that these pictures were intended to distract from the counting task. I instructed the participant that she must not look away or close her eyes, as the goal of the task was to measure the effects of visual distraction on attention. This slight deception was necessary to ensure that the participant focused on the counting task, since if she knew the real goal of the task, she might deliberately count slower to devote more attention to the picture sequences.

Both conditions began with a white 'READY' text displayed on a blank (neutral blue) screen for 500 ms, immediately followed by a 2000 ms blank screen during which an audio recording of a randomly-selected three-digit number played. At each prompt to count backward, the program displayed picture sequences of novel, yet identifiable, action scenarios. In total, four scenarios were presented, each consisting of a sequence of three pictures (see Appendix K for a list of scenarios). For example, the first picture might show a person putting an item in a box, the second show the person carrying a package, and the third show the person putting the package into a bin. All scenarios were designed such that the order of the images could be reversed and still make sense. If the participant remembered details semantically, but failed to remember the correct temporal sequence, I did not want the participant to be able to simply infer the order from semantic cues.

To control for the possibility that any effect observed was due to person-recognition in general and not self-recognition in particular, all scenarios featured a person. All photographs were taken against an undecorated backdrop, and featured exactly eight relatively simple items in the background—not including the items that played an active role in the scenario—each with a few distinctive features, such as a white teapot with a chrome handle and rusty spout. Photo editing software was used to remove unnecessary details and graphical artefacts. To match the presentation of the scenario in the HSR condition, the face of a high-profile figure (that is, a familiar yet not intimately known person) was digitally inserted over the face of the protagonist. I opted to use picture-sequences instead of short videos because it would have been too difficult to digitally insert the face into a video file. Each photograph in the sequence was presented for 2000 ms, with 200 ms intervals containing a blank screen (neutral blue).

In the HSR condition, the Backward Counting prompt was accompanied by the same pictures as in the LSR condition, except that instead of inserting a picture of a celebrity's face over the protagonist's face, I inserted a picture of the participant's face. In the interests of time, and so as not to interrupt testing, the insertion of the latter's face was done automatically using a custom Python script.

Studies have demonstrated that novelty increases the likelihood of episodic encoding (Tulving & Kroll, 1995; Tulving, Markowitsch, Craik, Habib, & Houle, 1996); I wanted to ensure that the LSR and HSR conditions were equally memorable in every respect other than self-relatedness. There was also a possible confound from the von Restorff effect, which is the effect of distinctive stimuli causing an item to be more memorable. Cohen and Carr (1975) observed the von Restorff effect with respect to faces, so there was good reason to control for it. I wanted to ensure that the effect being measured was due specifically to *self*-reference and not

merely to a sense of novelty. Therefore, I made stimuli in both conditions distinctive, with the only difference being that HSR pictures included the participant. Non-self pictures featured celebrities Justin Timberlake and Lindsay Lohan (who were culturally relevant at the time this pilot was being run), as well as former U.S. President George W. Bush, Jr. and current U.S. President Barack Obama, in keeping with Symons and Johnson's (1997) observation that descriptions of merely familiar people do not engender SREs; I wanted to ensure that, while distinctive, the stimuli would not trigger an SRE. Since it was unlikely that any of the participants had a personal relationship with these high-profile figures, pictures of such figures were ideal candidates.

Episodic free recall – pilot version.

This task tested the ability of participants to retrieve episodic memories about the tasks. Five minutes after the BC/TS task, I asked participants to complete a questionnaire (see Appendix G) designed to test their ability to freely recall details of the Temporal Sequence task in the order of presentation. The questionnaire consisted of eight questions, such as what the person was doing in the sequence, what happened in each frame, and what items could be seen in the room depicted. The questionnaire also asked the participant to describe each item in as much detail as possible. The same set of questions was asked for each of the four scenarios presented earlier.

I scored participants on the number of items that they could recall, as well as the number of details for each item, and then added these together to generate a level-of-detail score. There was no determined upper limit to this score, as I thought it possible that participants would remember numerous details about the items. I attempted to keep the items simple enough that there would only be a few distinctive features for each item. I also scored participants on

'sequence accuracy'—that is, how well they could recall the temporal order of events, because episodic memories encode—in addition to the *what, where,* and *when* of events—the order of events (Gardiner, 2001). As the picture sequences consisted of only three frames, scoring consisted of a simple pass/fail. If the participant remembered any of the frames out of sequence, she was given a score of '0', otherwise she was given a score of '1.'

Method – Follow-up study

After running a preliminary analysis partway through the study, I became concerned that poor performance on the episodic recall task might be indicative of a general difficulty with the task itself, independent of any effects of distraction by the Tetris game. To determine whether or not this was the case, I ran a small number of participants on a special version of the study, identical to the existing version only without the Tetris game as a distraction.

Participants.

Ten participants (six female and four male) between the ages of 17 and 22 ($M = 19$, $SD = 1.7$) were recruited from Carleton University's undergraduate psychology pool. Informed consent was obtained from all participants. Nine participants spoke fluent English. One participant spoke a dialect of Chinese fluently (but did not identify which one) and claimed not to be fluent in English, but was tested anyway and did not seem to have difficulty with the tasks. One participant was fluently bilingual in English and French, one participant was fluently bilingual in English and Arabic, and two spoke some French but not fluently. I assigned five participants to each condition (LSR and HSR, just as in the regular version of the study).

Design.

All participants were tested individually by a single experimenter (the author). Participants completed six tasks in one session, lasting approximately 45 minutes to an hour. The

design of the study was identical to the design of the full study, with the exception that the dual-task T/ATS was replaced by the lone Auditory Temporal Sequence task.

Materials and procedure.

Auditory temporal sequence.

This task was simply a single-task version of the Tetris/Auditory Temporal Sequence task described above. I made every attempt to vary the procedure as little as possible from the regular version of the study. Instead of telling the participant that she was going to play a game of Tetris and that the narrative was meant as a distraction, I simply asked the participant to listen to a story. As before, I did not tell the participant that she would be tested on the details of the story, though it is likely that many participants surmised that they would be expected to remember the details. If the participant asked whether or not she would have to remember the story, I simply told her that I couldn't reveal anything about the purpose of the task. The purpose was revealed at the end of the session and a "consent to the use of data" form administered to ensure that the participant understood why this slight deception was necessary, and still allowed her data be used.

Results – Full Study

Self-referential adjectives/Abstract visual distraction.

I ran a paired-samples t-test on the number of self-related words recalled and number of other-related words recalled. As expected, I confirmed the effect of self-reference on semantic recall in the Self-referential Adjectives/Abstract Visual Distraction task, recalling more self-related words than other-related words, $t(39) = 5.55$, $p < .001$. This result is, however, very well-established in the literature, as reported in the discussion of Symons & Johnson's (1997) meta-analysis in *Research on Indicators of SAS Consciousness*.

Table 1. *Descriptive statistics for all measures reported.*

				Skewness		Kurtosis	
Measure	Range[25]	Mean	SD	Statistic	SE	Statistic	SE
Age (years)	21	21	4	2.62	0.37	8.73	0.73
Cognitive Failures Questionnaire score (out of 100)	50	44	13	-0.33	0.37	-0.43	0.73
Visual attentiveness score (out of 6)	6	5	2	-0.89	0.37	0.39	0.73
Number of self-related words recalled	16	6	3	0.92	0.37	2.10	0.73
Number of other-related words recalled	7	3	2	0.29	0.37	-0.53	0.73
Total score across all games played	29	12	8	0.28	0.37	-1.03	0.73

[25] Expressed as the difference between highest and lowest value.

Epsiodic recall - level of correct detail	17	7	5	0.42	0.37	-0.79	0.73
Episodic recall - total level of detail	27	12	6	0.33	0.37	-0.48	0.73
Episodic recall - number of answers (out of 16)	14	10	4	-0.06	0.37	-1.07	0.73
Episodic recall - level of detail for character-related questions	5	2	1	0.14	0.37	-0.75	0.73
Confidence score for correctly remembered words (out of 120)	76	83	22	-0.12	0.37	-1.34	0.73
Remember/Know - Number of correct novel words	20	13	6	-0.99	0.37	-0.36	0.73
Remember/Know - Number of correct familiar words	13	17	3	-1.36	0.37	1.21	0.73
Remember/Know - Total number of words recalled	20	11	4	0.28	0.37	0.94	0.73
Remember/Know - Number of words recalled correctly with high confidence	37	24	10	-0.26	0.37	-0.88	0.73
Remember/Know - Number of words recalled correctly with moderate confidence	35	6	6	3.19	0.37	14.00	0.73
Remember/Know - Number of words recalled correctly with low confidence	7	1	2	1.90	0.37	2.81	0.73

Remember/Know - Total number of remembered words	107	78	23	-0.67	0.37	0.61	0.73
Remember/Know - Total number of known words	103	26	18	1.89	0.37	7.00	0.73
Remember/Know - Total number of guessed words	29	7	7	1.56	0.37	2.16	0.73
Remember/Know - Number of missed novel words	10	3	2	1.11	0.37	0.69	0.73
Remember/Know - Number of missed familiar words	18	14	5	-0.90	0.37	-0.10	0.73
Remember/Know - Total confidence score (out of 240)	78	208	20	-0.51	0.37	-0.14	0.73

Note: n = 40 for all measures

Visual attentiveness/Sustained inattentional blindness.

Participants scored highly in visual attentiveness (see Table 1). This suggests that any lack of effect observed between groups is unlikely to be a result of variability in visual attentiveness, although this was still controlled for in the relevant analyses (described below).

Tetris/Auditory temporal sequence/Episodic free recall.

Twenty-nine participants (73%) were familiar with the game of Tetris while eight were not (20%). Data for the remaining two participants (5%) were missing. Missing data were counted as unfamiliar for the purposes of analysis.

Recall Hypothesis 1: *The encoding of episodic memories is possible only when one is conscious of SAS*. To test this hypothesis, I conducted an independent-samples t-test comparing total level of correct detail between the HSR and LSR groups, and found no significant effect of other-reference, $t(38) = -0.031$, $p = .98$. Nor did I find a significant effect of distraction on total

level of detail (both correct and incorrect detail), $t(38) = -0.63$, $p = .54$, or on the number of questions answered, $t(38) = 0.92$, $p = .37$. In order to confirm that an effect was not present, and not simply being obscured by the participant's distractibility (as measured by the CFQ) or visual attention (as measured by the VA/SIB), I conducted a between-subjects ANCOVA comparing HSR and LSR conditions and still found no significant main effect of other-reference on level of correct detail, $F(1, 36) = 2.35$, $p = .75$. There was also no significant effect of other-reference on the total level of detail, $F(1, 36) = 0.23$, $p = .64$, nor on the number of answers given, $F(1, 36) = 0.66$, $p = .42$. Since I was not blind to the hypotheses being tested, I employed another rater to determine the reliability of the detail measures. I observed excellent inter-rater reliability on both total level of detail (ICC = .93) and level of correct detail (ICC = .92).

As expected, in the HSR condition there was a positive correlation between level of correct detail and confidence reported for remembered words, after controlling for distractibility and visual attention ($r = 0.54$, $n = 19$, $p = .03$), however this relationship was not significant after running a Bonferroni correction. In addition, a finer grained analysis turned up no significant difference between HSR and LSR groups for character-related level of detail (details pertaining to the protagonist of the narrative), $t(38) = -1.06$, $p = .30$.

In the LSR condition, as predicted, there was no correlation between level of correct detail and degree of confidence reported for remembered words in the LSR condition ($r = -.21$, $n = 21$, $p = .39$). These results suggest that my purported measure of episodic memory is not measuring the same phenomenon as Tulving's Remember/know task.

Unexpectedly, participants in the HSR condition identified the protagonist as someone else significantly more often than they identified as self ($\chi^2 = 4.3$, $p = .04$). As I expected, in the LSR condition participants correctly identified the protagonist more often than they did not ($\chi^2 =$

5.8, $p = .02$). For the purposes of analysis, I considered the protagonist correctly identified if the participant used the name "Stephen Harper," "Stephen," "Harper," or "the Prime Minister" in the LSR condition, or "me" or similar self-referential term in the HSR condition.

Remember/Know.

Participants scored moderately well on episodic memory recall ability, as measured by the total confidence score for correctly remembered words (see Table 1). No participant scored fewer than 43 out of 120 though the highest score recorded was close to the maximum possible, with 119 out of 120.

Participants recalled most words (both correct and incorrect) with a high degree of confidence. Participants recalled few words with a moderate level of confidence, and even fewer words with the lowest degree of confidence (see Table 1).

In terms of number of correct words remembered with varying degrees of confidence, participants recalled only moderate to few correct words to the highest degree of confidence. Participants recalled very few correct words with a moderate degree of confidence, and almost no correct words with the lowest degree of confidence (see Table 1). For the total confidence in remembering all words (out of 240), including words not in the study list, participants scored highly (see Table 1).

I ran paired-samples t-tests comparing number of familiar words recalled and number of novel words recalled. Participants tended to correctly remember familiar words (words presented since the beginning of the task) significantly more often than novel words (words that appeared only in the last set of trials), $t(39) = -3.59$, $p = .001$. However, participants tended to miss familiar words significantly less frequently than they missed novel words, $t(39) = -14.36$, $p < .001$.

Cognitive failures questionnaire.

Participants scored moderately in cognitive failures (see Table 1). No participant scored higher than 65 out of 100 or lower than 15.

Results – Follow-up Study

Auditory temporal sequence/Episodic free recall.

Since there were only ten participants in the follow-up study and forty in the main study, I compared the participants in the follow-up to a random selection of ten participants from the main study. As expected, Participants remembered significantly more correct details when administered the narrative by itself versus the narrative with the Tetris distraction, $t(18) = -4.04$, $p = .001$. Participants also reported significantly more detail (ignoring veracity) when presented with the narrative by itself than with the distraction $t(18) = -4.16$, $p = .001$.

Discussion

If it had been significant, the positive correlation (as reported earlier, $r = 0.54$, $n = 19$, $p = .03$) between level of detail recalled and episodic recall ability (as measured by the confidence reported for remembered words in the Remember/Know task) in the HSR condition would have supported my first hypothesis, that episodic memory encoding was dependent upon SAS consciousness, but only for participants with strong episodic encoding ability. Regardless of outcome, I would not have been able to conclude that there was a causal relationship. It is possible that episodic encoding ability depends upon SAS consciousness, or that both faculties have a common cause yet to be identified. Although the result was not statistically significant, it was close enough that it would be worth investigating in a replication of this study.

It is also possible that the Tetris task was not demanding enough to distract participants to the level needed to reduce SAS consciousness. It could also mean that no reduction occurred. I

suspect that the latter is the case, though I cannot say so with certainty due to the high degree of variability between participants. With the existing variability, a power analysis revealed that the sample size would need to be increased to an impractical size ($N > 1000$) to see any effect. Where possible, I controlled for the most obvious sources of variability but there are likely factors that I am unaware of or that cannot be controlled for. It would not be worth replicating the study as-is and merely collect a much larger sample. Rather, future studies should focus on improving the design to greatly reduce the variability. In the remainder of the discussion I will focus on the weaknesses of the design and make specific suggestions on how to improve the study.

I suspect that the Tetris game may not be producing the intended effect in enough cases, because it does not hit the right balance between being too easy and too challenging. Future studies could try increasing the difficulty of the game, or using a different game.

Many of the participants commented that the episodic recall task was quite hard, raising the possibility that the task's difficulty was washing out any effect. In order to rule out task difficulty as a confounding factor, we ran 10 additional participants with the narrative alone (no Tetris game). As expected, participants performed much better without the distraction, suggesting that task difficulty is not an issue.

Several participants in the HSR condition identified the protagonist of the narrative as someone other than themselves. I suspect that this may have in part been due to confusion about the use of the second-person pronoun in that narrative. It is an unusual narrative construction and they may have been unsure how to conceive of the protagonist, causing some to assume the story was about an unspecified person. Swapping the second-person pronoun for a first-person

pronoun may also help, although the first-person pronoun is more likely to lead the listener to assume that the unidentified narrator is the protagonist.

It is of some interest to note that some women assumed the protagonist in the HSR condition to be male, even though there were no cues in the story as to the sex of the protagonist. This could be due to not understanding that the "you" was intended to refer to them, or it could be due to the narrator's (my) voice being male. Though I did not collect any data on this, future research could try recording the narrative with both a male and female narrator. The reason I did not elect to do so in my study was out of concern for consistency—I thought that having an additional version of each narrative could introduce unexpected confounds.

The questions asked in the narrative were mostly "what" questions. For a more fine-grained analysis, future narratives should contain more "when" and "where" questions, to allow comparisons between different kinds of episodic details.

Study II: Diachronic SAS Consciousness in Young Children

Hypotheses

2) *By the age of three-and-a-half years, typically-developing children are capable of SAS consciousness.* This hypothesis is motivated by general interest in the structure of SAS consciousness.
3) *Before the age of approximately five years, children have yet to fully integrate SAS consciousness with SAO consciousness, as evidenced by an inability to recognize previous images of themselves as themselves, despite having episodic memory of recent events.* This hypothesis is also motivated by general interest in the structure of SAS consciousness, but seeks to expand on this structure by exploring the ability to integrate SAS consciousness with SAO consciousness.

Introduction

There is sufficient evidence in the literature that children have consciousness of both SAS and SAO to some degree by the age of 18 months. As Povinelli, Landry, Theall, Clark, and Castille (1999) note, however, as late as their fourth year children show deficits of self-attribution over the passage of time. In other words, children at this age have difficulty understanding how to attribute past events to present instances of self, such as understanding that a sticker placed on their heads in the past has a relationship to the presence of a sticker on their heads in the present. It is also around this time that, on some accounts (see Vandekerckhove, 2009 and Picard *et al*., 2009, above), children begin to demonstrate a capacity for episodic memory. Peterson's (2002, see above) competing account claims that children demonstrate *some* episodic memory capability as early as two years, but even Peterson acknowledges that early episodic memories (before four years) are sporadic and brief in duration. The observations of all of these studies suggest that interesting developments in consciousness of self (both SAS and SAO) occur around the four year mark in typically-developing children. Unfortunately, I cannot conclude anything from simple correlations among these two abilities and age. Since the age range of interest is a period of rapid development in children, age is not a reliable indicator of overall development. In the study described below, I included a measure of receptive vocabulary (i.e., comprehension), which has been established as measure of verbal intelligence (see for example Campbell, Bell, & Keith, 2001) and as such is a more reliable developmental signpost than age.

Since there have been no studies with the explicit aim of testing for SAS consciousness at an early age, one component of this study explores hypothesis (2), that *by the age of three-and-a-half years, typically-developing children are capable of SAS consciousness*. For practical

reasons, I did not test children younger than about three years, so I could not measure the onset of SAS consciousness, which I expect occurs prior to first-person pronoun mastery (occurring at 18 months, on average). Due to the preliminary nature of this study, it was not possible to test this hypothesis fully, and the results presented below should be considered tentative, though they present interesting avenues for future research.

In Hypothesis 3, I claim that *before the age of approximately five years, children have yet to fully integrate SAS consciousness with SAO consciousness, as evidenced by an inability to recognize previous images of themselves as themselves, despite having episodic memory of recent events.* Put slightly different, by the age of five years, children develop the ability to recognize previous instances of SAO *as* previous instances of themselves.

The investigation of the third hypothesis concentrates on the development of the self-specific aspect of episodic memory encoding and recall. It is also intended to address some of the weaknesses of Povinelli's studies (see *Empirical Support for SAS/SAO*, above), and expand on our knowledge of the link between episodic memory and consciousness of self. Once again, the preliminary nature of the study renders my results tentative, though suggestive of promising future directions.

I ran a time-delayed self-recognition task similar to the one used by Povinelli, except that I used a still image instead of video due to privacy concerns.[26] Povinelli's task involved placing a sticker surreptitiously on the child's head during play, then showing her the picture/video of the sticker placement several minutes later. If the child reached for the sticker upon seeing the

[26] This was also due to the consideration that it would already be difficult to get children and their caregivers onto campus, and it would be easier to find willing participants if they were not concerned about video being recorded of their child.

image(s), this was taken as evidence of self-recognition extended into the past. Povinelli's study demonstrated that the same measure could be performed with video or still imagery.

The aim of my study is to test whether or not children who fail a past-self recognition task like Povinelli's nevertheless have an understanding of themselves as a single entity extending into the past. Povinelli's study tested only whether or not children could recognize past instances of themselves—in other words, past instances of SAO. If children who fail to recognize these earlier instances of self can nevertheless remember past events as events that happened *to* them, this would indicate that they have SAS consciousness extending into the past, but have yet to connect multiple representations of SAO (i.e., the set of previous instances of SAO) to SAS. I combined the past-self recognition test with a pair of tasks, one to evaluate a child's mastery of the first-person pronoun, and one to measure the child's capacity for episodic memory recall. This combination of tasks is an innovation on the tasks designed by Povinelli, which did not dissociate SAS from SAO, evaluate pronoun use systematically, or test for episodic memory ability.

Method

Participants.

In total, twenty-four children (11 female) between the ages of three and six years, or 39 and 83 months (M = 4.5 years or 54 months, SD = 1.0 years or 11.5 months) were tested. Six were excluded from analysis because they discovered the sticky note before the critical test, two were excluded because they were tested with early versions of stimuli and had some missing data (one of these was too old for the analysis, and was tested for the express purpose of testing stimuli), and two were excluded due to lack of responsiveness on several questions, either due to the refusal to answer or inability to understand the questions. All participants spoke English as a

first language. Nineteen were monolingual and three were bilingual, while two spoke some French but were not fluent.

No SES (socio-economic status) data were collected, nor did I record or control for ethnicity. Caregivers were also asked to participate in the session, however no data were collected about the caregivers. The remaining participants were divided into three age groups, five in the three-year-old group, five in the four-year-old group, and four in the five-year-old group. Due to the small number of participants in each group, it was not possible to run meaningful inferential analyses, as I discuss in *Results*.[27]

Design.

All participants were tested individually by a single experimenter (the author) in a child-friendly lab environment. Each session lasted approximately 45 minutes, not including a short warm-up period before the start of the tasks. The study contained six measures, listed in the order they were administered: the Caregiver Report on Event Memory and Pronoun Use, Present-self-recognition Task (Who Has the Picture), Episodic Encoding Event, Past-self-recognition Task, Episodic recall Task, and the Peabody Picture Vocabulary Test (PPVT-III, Dunn & Dunn, 1997). Due to the nature of the tasks, it was not possible to randomize the order of presentation. In some instances, PPVT data from a recent, unrelated study in the same lab were used instead of administering the task again (after securing required parental consent and approval from the ethics review board). I describe each of these measures in detail in the following section.

Materials and procedure.

I used a digital camera connected to a laptop to capture a sequence of still images of the participants (deleted at the end of the session) and presented one of these images (selected based

[27] Analysis of variance (ANOVA) relies on the assumption that each group being tested has five or more participants. This requirement was violated by one group, and the remainder barely met it.

on criteria described below) as a stimulus on the laptop computer screen later in the session. Other visual stimuli consisted of plastic toy versions of food and animals, and picture cards featuring cartoon drawings of easily-recognizable objects, described in the tasks below. A digital audio recorder was used during three of the tasks (Present-self-recognition, Past-self-recognition, and Episodic recall) to ensure accuracy of recording.

Prior to each session, after introducing myself to both participant and caregiver, I briefed the participant's caregiver on the nature of the study and informed her that some still pictures needed to be taken during the session (caregivers were also informed before signing up for the study that the nature of the research required that pictures be taken). To assure the caregiver of her child's safety, I informed her that the computer was not connected to the Internet, that the images were necessary only during the test and would be deleted at the end of the session. I also informed the caregiver that she could watch her child's activities on the lab monitor (a T.V. screen connected to the video cameras filming in the testing rooms). After briefing the caregiver, I asked her to fill out the *Caregiver Report on Event Memory and Pronoun Use* (Appendix B), which was collected during a break prior to the Episodic-recall Task, described below. Next, I asked the participant if she would like to play some games with me. I then engaged the participant in a brief period of warm-up play, consisting of a 'cup game', described below. This warm-up session was not timed but typically ran no more than two or three minutes. Aside from helping the participant feel comfortable prior to testing, the warm-up period was required to prepare for later tasks. First, it was crucial that the participant knew my name, as the Present-self-recognition Task would evaluate use of personal pronouns versus proper names. Second, it was during the warm-up when I placed a sticky note on the participant's head, necessary for the past-self-recognition task. In the pilot phase, participants were encouraged to wear foam visors,

and the sticky note was placed on the visor to prevent participants from feeling the sticky note placement. The visors turned out to be more of a hindrance than a help, as most participants refused to leave the visor in place until presentation of the stimulus. In practice, none of the participants were able to distinguish the sticky note placement from an ordinary pat on the head.

The warm-up game was a recreation of one used by Povinelli *et al*. (1996), and provided me an opportunity to surreptitiously place the sticky note on the participant's head. The goal of the game was to find a toy egg hidden beneath one of four cups. In Povinelli's version, participants were asked to look for stickers and then place the stickers on a piece of paper. To avoid confusion with the sticky note revealed later, I did not want to introduce stickers as stimuli. The game consisted of five trials. The set-up for each trial was as follows: behind a movable barrier (an opaque plastic clipboard) and out of view of the child, I placed four over-turned paper cups, identical in size and appearance, and placed the egg under a randomly chosen cup. I then lifted the barrier and asked the participant to find the egg by lifting the cups. During the first and second trial, after the participant found the egg, I gave her verbal feedback (e.g., "great job!") and a pat on the side of the head facing the camera. After setting up the third trial, I pressed a key to engage the camera, which automatically took a snapshot twice per second for one minute. During the third trial, I repeated the same procedure as the first two trials, but used the pat on the head as an opportunity to place a bright orange sticky note on the side of her head, over the hair. In the fourth and fifth trial, only verbal feedback was given, and the camera was not engaged. Children who discovered the sticky note for any reason prior to the Episodic-recall Task were excluded from the analysis.

Caregiver report on event memory and pronoun use.

Some studies (Lewis & Carmody, 2008, Lewis & Ramsay, 2004, and Stipek, Gralinski, & Kopp, 1990) have used maternal reports, consisting of asking mothers series of questions regarding their children's use of the first-person pronoun. Maternal or primary-caregiver reports are useful because they cover a much larger time-span and range of behaviour than a single lab session can capture. I used caregiver reports to collect information on how and how often the participant used the first-person pronoun, as well as information about how often the participant engaged in particular activities. The latter was used to inform the episodic memory recall task later in the session.

The set of questions were adapted from those used by Stipek *et al.* (1990), and are shown in Appendix B. The first set of questions asked the caregiver to provide examples of significant events in which the child had participated in three different time frames: within the past year but not within the last 6 months, within the past 6 months, and within the past week. Caregivers were also asked to specify whether or not the child had attended a birthday party in each of these time frames, and whether or not the child had her own birthday party in each of these time frames. This information was used in the later Episodic Recall task to ask the child about various past events she had experienced. Finally, caregivers were asked a series of multiple-choice questions about how often the child engaged in particular activities or exhibited certain behaviour. Caregivers were allowed to fill out the information while I began testing the participant. I collected the form during a break before the Episodic Recall task, and entered the data into the protocol during the break.

Present-self-recognition task (Who has the picture?).

This task was designed to investigate the second hypothesis, that children over the age of three-and-a-half possessed SAS consciousness in the present. The task was based on the expectation that proper use of the first-person pronoun was possible only if the child was aware of herself as a subject of experience, at least in the present. For the purposes of my study, I also treated self-pointing as indicative of SAS consciousness in the present. Many children did rely on self-pointing even though they were capable of first-person pronoun use.

To measure SAS consciousness in the present, I used a simplified version of the first task employed by Lee *et al.* (1994) (see the previous section for more on the details of that study). I used eight pairs of white picture cards, each 32 x 26 cm with a 10 x 7 cm image pasted in the center of the card—the same dimensions used by Lee *et al.* (1994). Images on the cards were coloured cartoon-style line drawings of familiar objects, including animals, beach-related items, and party-related items (see Appendix C).

To administer the task, I took a seat adjacent to the participant, and began the audio recording. In each trial, I placed a pair of picture cards on a table, one card in front of the participant and the other in front of me. I then asked the participant, "who has the *x*?" and "who has the *y*?" where *x* and *y* were the objects displayed on the participant's card and my card, respectively. On some trials, *x* and *y* were reversed to rule out the possibility that participants would simply memorize the order.

The procedure for the task was as follows:

1) I displayed each card in the set to the participant, and asked her to identify the object on the card. This step was just to ensure that the participant knew the names of the objects on the cards. For each card, I recorded whether or not the participant knew the name of

the object. If the participant did not use the correct name of the object, I recorded whatever the participant called it.

2) Sitting adjacent to the participant, I placed card *x* face-up in front of the participant and card *y* face-up in front of me.
3) I asked the participant "Who has the *x*?" where *x* was the name of the object—or, the name that the participant called it instead, as the case may be—then recorded the answer (e.g., I/me/you/other). On some trials, step 3 and 4 were swapped as per the reasoning described above.
4) I asked the participant "Who has the *y*?" where *y* was the name of the object—or, the name that the participant called it instead, as the case may be—then recorded the answer (e.g., I/me/you/other).
5) I repeated steps 2-4 seven more times with the next pair of picture-cards, for a total of eight trials.

I scored participants based on how consistently they used (or did not use) "I" and "you" pronouns: consistently correct (7 or higher out of 8), consistently reversed (1 or lower out of 8), inconsistent (2 to 6 out of 8), or failed to respond (no data recorded). I recorded if the participant used proper names instead of pronouns, and if so, how frequently.

Episodic encoding event.

At this point in the session, I presented a memorable event to the participant. In the context of a "picnic game" I presented the participant with a series of plastic models of foods that she was likely to have strong preferences for and non-food items that she was likely to have strong preferences against. Eight plastic models were used: four of appetizing foods (pizza slice, hot dog, strawberry and pie slice) and four of unappetizing items (snake, lizard, rat and spider). I

provided a small basket and garbage can in which the participant placed items during the course of the task.

At the start of the task, I informed the participant that we were about to play a picnic game, and that none of the items I was about to show her was real. I placed the basket and the small garbage can (a blue bin) in front of the participant and told her to put the items that she would want to bring on a picnic into the basket, and put the items she would not want to bring into the garbage can. I brought out a green bin containing the models and removed a model, placing it in front of the participant. When the participant placed the model in either the basket or garbage can, I recorded where the item was placed. I then took out the next item and repeated the process seven times, each time with a new item. For the purposes of this task it did not matter whether or not the participant's own preferences and dislikes matched the items, since the point of the task was simply to provide memorable experiences. For example, some children disliked strawberries while others were partial to the spider.

This kind of event seemed ideal for encoding episodic memory because children were likely to take a personal interest in the stimuli, and I expected that children would have a moderate emotional reaction to the models. As observed in adults by Conway (2005), episodic-autobiographical memories are closely associated with visual mental imagery and emotional experiences. The task was expected to elicit a response strong enough for episodic encoding but not so strong as to raise ethical concerns, in the case of possible negative reactions. Children were also likely to have vivid mental imagery of items that they enjoyed or strongly disliked. As a precaution, I informed the caregiver at the beginning of the session that a few "scary" items like snakes and spiders might be used as props, but that they would not be presented in a scary or

frightening manner, and that the child would be informed ahead of time that the items were not real.

Past-self-recognition.

At the beginning of the task, I told the participant that I was going to show her a picture on the computer screen and ask her some questions about it, and started the audio recorder. After finding an appropriate image (one that clearly showed me placing the sticky note) I presented the participant with one of the snapshots taken earlier, displayed at full-screen on the computer. As in standard mirror self-recognition studies, the participant was monitored to see whether or not she reached for her head upon seeing herself in the image. If the participant did not reach for the sticky note, I asked her a series of questions designed to provide insight into her reasons for not reaching for the sticky note; these were the same questions described in Povinelli *et al*. (1996) plus a few new ones (see Appendix D for a complete list in order of presentation). I ran through the list of questions until the participant reached for the note, clearly indicated that the note was on her head, or exhausted the questions. If the participant did not discover the note, even after prompting, I removed the note and said 'here it is, it was on your head!' I also kept track of what words or gestures the participant used to describe the person (herself) in the image—for example, 'me', 'Patty', 'the girl' or 'her'—as these would be informative about whether or not the participant could integrate earlier instances of SAO with SAS.

While Povinelli *et al*. (1996) simply recorded the answers to these questions, I generated a total score based on how much prompting the participants required. I scored the participant by giving her full points if she reached up for the note immediately upon seeing the image, subtracting one point for each prompt, to a minimum of zero points if the question list was exhausted.

Immediately after the note was removed, I asked each participant a number of questions—depending on the questions (if any) already asked—about the event (also listed in Appendix D), such as why she reached for the note and who she saw in the picture. These questions were designed to provide information about the participant's reasons for reaching up for the note, and were similar to the questions that may have been asked prior to note-removal. The answers provided insight into whether or not children who reached for the sticky note conceived of themselves as extending into the past. I recorded the answers to these questions on the scoring sheet, using the audio record to ensure accuracy. For each question, I gave the participant three points for answering the question correctly without further prompting,[28] two points for answering correctly after a single prompt, one point for answering correctly after more than one prompt, and zero for answering incorrectly either immediately or after further prompting. A refusal to answer was coded as 'no response' and given a score of zero, however one case where the child refused to respond was excluded from analysis. The total score was calculated as the mean score of all responses.

Episodic recall.

In this task, I measured the participant's ability to recall episodic features of the encoding event. In the first part of this task, I asked the participant a series of questions about *where, when*, and *in what order* the picnic event took place (see Appendix E for a preliminary list), and whether or not she liked the items represented. The *in what order* question was of particular importance in light of Picard, Reffuveille, Eustache, and Piolino's (2009) observation that recall of proper temporal—and causal—sequence is a key indicator of episodic memory. In the second part of the task, I asked the participant a similar series of questions about events—identified

[28] A question counted as correct if the child said "me," self-pointed, used her own name, or otherwise indicated that the person in the picture was her.

earlier from the *Caregiver Report*—taking place in more remote time-frames (up to one year ago). I first asked the participant about an event (either a birthday party or other event) that happened up to a year ago but no more recently than six months ago, using the 'pre-test' questions specified in Appendix E. Using the same questions, I then asked the participant about an event that happened within the past six months (specified in the *Caregiver Report*), and an event that happened in the last week, for a total of three events.

I scored participants based on how much detail they provided on each question, as well as the total score on all questions pertaining to a single event, generating a level-of-detail score for each event. The score did not measure accuracy of recall, as I was simply interested in the amount of detail, not its veracity. As a result, even fantastical or clearly incorrect details counted toward the total. I did not evaluate the veracity of the details because the key feature of episodic memory, remembering from the standpoint of the experience, does not depend on whether or not the memory is accurate. Had it been a trivial matter to include a measure of veracity, I would have done so, however there were numerous practical obstacles to collecting reliable information about the participant's past experiences. For example, I could have asked the parent to verify the participant's responses, however this would only give a general sense of the veracity (and would be entirely dependent on the caregiver's own memory ability). Additionally, the caregiver may not have witnessed the event that the participant had attended.

I recorded the answers in audio format and transcribed them into digital form before coding. Level of detail was calculated using the following formula:

- Add one point for each unique feature (either a noun or noun phrase not repeated in this or any other answer).
- Add one point for each unique adjective (the same adjective used in two places

describing two different features would count as two unique adjectives).

- "I don't know," "nothing else," or similar negative response does not count as a detail.
- A simple "yes" or "no" in response to a prompt does not count as a detail.
- If a detail is merely a repetition of what the experimenter said, it does not count as a detail.
- If the participant said something marked as "[unclear]" in the transcript, it counts as a detail, unless it is being used as an adjective to describe a noun, as for example "[unclear] potato."
- Add one point if the participant names or identifies people in their response, but do not count each person as a unique detail.

Experimenter prompts were transcribed enclosed in square brackets and did not count toward the total, but details provided by participants in response to those prompts did count. The transcribed items marked "[unclear]" counted as details for the sake of consistency, though I did not count them if used as adjectives because I was less certain that the modifiers were meaningful in those instances. I added only one point if the participant identified any number of people because I didn't want to bias the score in favour of participants who just happened to be with a large number of people during the event in question. While participants who can remember more people should be given a higher detail score, there was no way to determine after the fact whether or not the response is due to memory ability, or simply a reflection of the fact that there happened to be a larger number of people at the event, compared to events attended by other participants.

Peabody picture vocabulary test (PPVT-III).

Finally, I administered the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), to provide an additional developmental signpost. The PPVT-III is a standardized test of receptive vocabulary and verbal intelligence, which is a more reliable indicator of general cognitive development than age. The results of this test were used to control for the effects of language and general cognitive development, since these factors introduce a substantial amount of variability in the abilities of children this early in development.

I instructed the participant that I was going to show her some pictures and ask her to point to the one I named. I began the task by showing a couple of training pages in sequence. Each page contained four panels depicting a different object or activity. I would then say, for example, "can you put your finger on *ball*?" and record which panel the child pointed to. If the point was ambiguous in any way, I would repeat the question until the child placed her finger on a panel. If the child pointed to the correct items in the training pages, I moved immediately onto the testing pages, otherwise I would say "try again, can you put your finger on *X*?" until the participant pointed to the correct item.

In the testing trials I would proceed as stated above, except that I would simply record what panel the child pointed to without giving any feedback, whether or not the choice was correct. The numbered pages were organized by order of difficulty into sets of twelve, and the sets themselves organized into age groups. As per the task's administration guidelines, if a participant got eight or more items incorrect in a single set, I would end the task, recording the numerical index of the final item as the 'ceiling' item. Typically I would begin with the first set in the age group ahead of the participant's own, in order to expedite the task (out of concern that children would become bored and abandon the session). If the child made two or more errors on

the set presented, I would present the previous set and work backwards until the child made one or fewer errors on a set (to establish 'basal', as set out in the administration guidelines). I would then proceed as normal with the next highest incomplete set.

I assigned the participant a raw score calculated by subtracting the number of errors (pages incorrect) from the numerical index of the ceiling item.[29] The raw score was used in the analysis. Several of the participants were not administered the PPVT during the session because they had recently been tested on the PPVT in a previous, unrelated study in the same lab. Since the PPVT is a standardized test, and the child's receptive vocabulary was unlikely to change in the brief period of time between studies, this was a useful strategy to streamline the administration time of the study. Further, a child's score on this test is considered to be stable for at least six months. Likewise, PPVT data collected during my study were used in subsequent studies by other researchers.

Debriefing.

At the end of the session I debriefed the caregiver and gave the participant a book and sheet of stickers as a 'thank-you' for participating (even if the child did not complete all of the tasks).

Results

Participants.

Fourteen children (7 female) between the ages of three and five years, or 39 and 71 months (M = 4.6 years or 55 months, SD = 1.0 years or 11.5 months) were tested and included in the analysis. Among these, receptive vocabulary (in raw score) was normally distributed (M = 84, SD = 19). One of the 14 participants in the analysis was missing data for when she reached

[29] Items are numbered in order of presentation, so the difference between the index and number of errors is how many items the participant correctly identified.

for the sticky note, as she refused to answer any of the questions regarding the photograph (hence the smaller *n* for self-recognition score in Table 2). Data for the other tasks were complete, however, so her data were included in the analysis, where applicable.

Table 2. *Descriptive statistics for all measures reported.*

Measure	N	Range	Mean	SD	Skewness		Kurtosis	
					Statistic	SE	Statistic	SE
Age in months	14	32	55	12	0.09	0.60	-1.26	1.15
PPVT raw score	14	68	84	19	0.62	0.60	0.04	1.15
How many pictures the child correctly identified (Who Has the Picture?)	14	1	16	0	-1.57	0.60	0.50	1.15
Number of errors identifying who had the picture	14	8	1	3	2.41	0.60	4.65	1.15
How many times the child used a pointing gesture *instead of* a pronoun or proper name	14	16	3	6	1.93	0.60	2.38	1.15
How many times the child used a pronoun *or* pointed	14	16	13	6	-1.68	0.60	1.65	1.15
How many times the child used a proper name (out of 16)	14	6	0	2	3.74	0.60	14.00	1.15
Self-recognition score	13	5	3	2	-0.43	0.62	-0.44	1.19
Total level of detail (LoD) across all events	14	20	20	6	0.77	0.60	-0.80	1.15
LoD for events occurring up to one week previously	14	12	7	3	-0.62	0.60	3.32	1.15
LoD for events occurring up to six months previously	14	8	6	3	0.43	0.60	-1.55	1.15

LoD for events occurring up to a year previously	14	11	8	4	0.63	0.60	-0.89	1.15
Caregiver Report score (out of 132)	14	48	106	12	-0.17	0.60	0.74	1.15

As mentioned previously, sample size was severely limited due to difficulties in recruiting participants. First, due to requirements imposed by the nature of the tasks, it was not possible to run the study in a daycare environment, which has proven to be most conducive to obtaining child data in a timely manner. The requirement of running the study in a campus environment creates a much greater imposition on the caregiver than simply signing a form sent home by the daycare, especially since I was asking for some participation on the part of the caregiver. Secondly, the requirement that children be recorded in both auditory and visual form is likely to have dissuaded some caregivers out of safety and privacy concerns.

My analysis is limited to descriptive statistics and correlations, as more advanced statistical techniques would not be reliable. All correlational analyses should be considered tentative, given the small sample and limited statistical power. I include them here only to suggest where future research might focus its efforts.

Caregiver report on event memory and pronoun use.

There was no significant correlation between episodic recall and caregiver-reported episodic memory ability, after controlling for age and receptive vocabulary (see Table 4). This suggests that my measure of episodic memory ability is not measuring the same phenomenon.

Present-self-recognition task (Who has the picture?).

Almost all participants correctly named all 16 pictures (see Table 2). This was unsurprising, as the stimuli were selected such that they would be easily recognizable by the youngest participants.

As expected, nearly all participants performed at ceiling on the pronoun assignment measure (making no errors), using the proper pronouns or pointing at the correct person when asked who had the picture (see Table 2). As a result, the measure would not have been meaningful in comparison with any other measures. It did, however, establish that most of the children could at least recognize themselves in the present, confirming hypothesis (2). Given the small number of participants, we should not assume that the sample is representative of the population.

Table 3. *Bivariate correlations for all measures reported.*

	Measure		2	3	4	5	6
1	Age (mos) (*n* = 14)	*r*	**.70***	-.11	.14	.13	.04
		p	.005	.70	.63	0.67	.89
2	PPVT raw score (*n* = 14)	*r*		.23	.20	.34	.23
		p		.42	.48	.30	.43
3	Caregiver Report score (*n* = 14)	*r*			.002	.61	.13
		p			.99	.03	.66
4	Pronoun use errors (*n* = 14)	*r*				.12	-.03
		p				.72	.91
5	Self-recognition score (*n* = 13)	*r*					.19
		p					.53
6	Total level of detail (LoD) for all events (*n* = 14)	*r*					--
		p					

*. Correlation is significant at the 0.01 level (2-tailed).

Only two children made mistakes, but they consistently reversed the answers (e.g., pointing at me or saying “you” when they were holding the picture in question) and had no

trouble using the first- and second-person pronouns throughout the remainder of the session, leading me to believe that they were either intentionally doing the reverse of what I was asking, or simply misunderstood the task.

Only four children used pointing gestures instead of using a pronoun or proper name, only two of whom used the gesture exclusively while the others used them sporadically. Most children used a pronoun and almost no children used a proper name in place of a pronoun or pointing gesture (see Table 2).

Table 4. *Partial correlations controlling for age and receptive vocabulary.*

	Measure		2	3	4
1	Caregiver Report score (*df* = 10)	*r*	-.05	.54	.01
		p	.87	.09	.97
2	Pronoun use errors (*df* = 10)	*r*		.08	-.08
		p		.82	.80
3	Self-recognition score (*df* = 9)	*r*			.06
		p			.87
4	Total level of detail (LoD) for all events (*df* = 10)	*r*			--
		p			

Note: partial correlations controlling for age (in months) and receptive vocabulary (PPVT-III).

Past-self-recognition/Episodic recall.

Hypothesis (3) remains unresolved, as there was no significant correlation between recognition of self-image and episodic recall, after controlling for receptive vocabulary and age (see Table 4). Self-recognition was measured on a scale from zero to five, where zero indicated no recognition and five indicated that the child reached immediately upon seeing the image. Each score below five indicated how many times the participant was prompted before reaching

(Score = 5 - $N_{prompts}$). If the participant still didn't reach after the fourth prompt, I stopped prompting and assigned a score of zero. On average, participants required about two prompts before reaching for the sticky note (see Table 2 – note that the number reported is [5 – $N_{prompts}$], not the number of prompts). However, this lack of finding is constrained by the small sample size.

I found a non-significant positive correlation between episodic recall and receptive vocabulary—as measured by the PPVT. For the measure of episodic recall of events taking place before the session, children were able to recall 20 unique details, on average, across all answers, (see Table 2). A one-way within-subjects ANOVA comparing the detail scores in the three different sets of event questions revealed no significant difference among the three sets, $F(2, 26) = 1.26$, $p = .30$. This suggests that children's episodic memories for events occurring up to a year previously were no worse than their memories for events occurring within the last six months or within the last week.

Discussion

I expected that I would not find a significant correlation between episodic recall ability and competence in the past-self-recognition test, since I expected episodic recall to depend on SAS consciousness, but not SAO consciousness. However, the absence of evidence cannot be taken as evidence to the contrary, especially given the small sample size.

I expected that children who had difficulty recognizing previous images of themselves *as* themselves would nevertheless have episodic memory of recent events. The lack of correlation between self-recognition and episodic recall was not surprising given hypothesis (2), since if episodic recall was dependent on SAS consciousness, and not SAO consciousness, I expected the requisite faculties to already be in place by age three. Nevertheless, I thought that I might

observe a correlation due to a developing understanding of SAO during this period. I also expected that low episodic-recall scores would correlate with low scores on the *Caregiver Report*, as these were both supposed to measure episodic recall ability. The lack of a relationship could mean that the measure of episodic recall I developed was not actually measuring episodic recall, or, less likely, that the *Caregiver Report* is not a reliable indicator of episodic recall ability. I reserve judgment on either conclusion until such time as the same measures can be performed with a much larger sample.

Though I observed a positive correlation between episodic recall and receptive vocabulary (as a proxy for developmental stage), it was not significant. This suggests that episodic recall ability does not significantly improve between the ages of three and five. I claimed that episodic recall depends on SAS consciousness, not SAO consciousness, and thus should be in place before the age of three. While my results appear to support this claim, more study is needed to confidently assert that there is a lack of improvement in episodic recall ability during this period of development. Additionally, the lack of a relationship between the *Caregiver Report* and my episodic recall measure lends some caution about asserting that claim prematurely.

While my results are not sufficient to fully address my last two research questions, they do demonstrate the promise of experimental methods in addressing those questions. Both the past-self-recognition task and episodic recall task provided variability among participants, although it remains to be seen what exactly the episodic recall task was measuring.

Conclusions and Future Directions

I posed the following questions at the beginning of the *Empirical Studies* section:

1) Does the encoding of episodic memories depend on SAS consciousness?

2) By what age are typically-developing children capable of recognizing themselves as SAS?
3) When does SAS consciousness become fully integrated with SAO consciousness?

I have addressed each of these questions in the form of specific hypotheses in the previous chapter. Here, I discuss in more general terms how the outcome of these studies informs these questions, and where future research in this area should focus its attention. Due to the inconclusiveness of the results, the emphasis is on the latter.

While the results of both studies were largely inconclusive, there are a few points that warrant further discussion. With regards to the first study, future studies could recast the first hypothesis in terms of the relationship between episodic memory and consciousness in general. It seems likely that if it is possible to be conscious without being SAS conscious, or with reduced SAS consciousness, the effect is so subtle as to be inaccessible to the kind of coarse-grained methods employed in this study. It is unclear how the existing methods could be made finer-grained without extensive pilot studies. The purported link between self-awareness and episodic memory is often taken for granted in the psychological literature, however based on my experiences it is far from obvious that this relationship exists, let alone what the nature of the relationship is. Though my results were inconclusive, the fact that they failed to verify a commonly held assumption should at the very least indicate that more study is needed to confidently assert that the relationship exists.

The developmental study, though preliminary, suggests the possibility that Povinelli's interpretation of his results is incorrect. As I explained in the discussion of his work, the possession of episodic autobiographical memory during the age range in question would count against his interpretation. I stated earlier that the SAS/SAO distinction would allow us to make

sense of the delayed self-image as a failure of integration between SAS and SAO. I was unable to find any discernable pattern to SAS/SAO integration, however I think it would be worthwhile to repeat the study with a much larger sample size if only to more confidently assert that there is no pattern.

My study focused exclusively on the brief delay condition, whereas Povinelli compared a brief delay to an extended delay of about a week. Due to laboratory constraints, it was unrealistic for me to run the same comparison, so my analysis cannot be used to comment on the possibility that SAS/SAO integration varies with the amount of time elapsed, such that older children would have a more robust sense of self over longer periods of time than younger children. Future studies should contrast the brief delay with a longer delay.

There are better developmental milestones than age, and previous research mostly characterized the onset of self-consciousness in terms of age. My study made use of a measure of receptive vocabulary as a proxy for developmental stage; however this measure would not be appropriate for children at 18 months or earlier.

In sum, future research on the development of SAS/SAO will require a much larger sample size; I estimate that a minimum of 60 children, with a 2-year-old group and at least 20 in each age group, would be sufficient. The study should also take place over at least two sessions to compare brief versus extended delays and use non-linguistic measures of episodic memory. With the exception of the latter, few methodological changes would need to be made. I would like to note, however, that based on my experiences in trying to recruit caregiver/child pairs and requiring them to come onto campus, it would be worthwhile to explore ways to conduct the study in the context of a daycare or other venue that does not place as large an imposition on the caregiver(s). This imposition largely stems from the use of the questionnaire on pronoun use and

episodic recall, so a task that measures use of the first-person pronoun but does not rely on caregiver reports would be ideal. Alternatively, the study could omit the pronoun use measure entirely and focus exclusively on episodic recall, though caregiver input would still be required to determine what past events to ask questions about. In a daycare environment, it might be possible to coordinate with daycare employees to determine when special events took place in the daycare, and ask the children about those events. This assumes, however, that the other difficulties with administering the study in a daycare environment could be appropriately dealt with, including for example the need to record audio and photographs of participants or the requirement of having a quiet, relatively distraction-free space to conduct the session.

I remarked at the outset of this work that research on SAS consciousness has been minimal, in part because few of the research questions asked by empirical researchers have been posed in light of the SAS/SAO distinction. My experience in attempting to study SAS consciousness has at times been one of great frustration. My attempt at manipulating attention to explore the relationship between SAS consciousness and episodic memory or to look for developmental markers has turned up dry, which is not to imply that we should abandon the search along those avenues. The state of research on this topic is far too preliminary to declare these avenues fruitless, although based on my own experiences and intuition I would recommend exploring other dimensions. In particular, it would be worthwhile to develop a deeper understanding of the mechanisms of episodic memory in general before tackling the question of how self-consciousness may (or may not) be related. It is clear, however, that we cannot take that relationship for granted.

References

Addis, D. R., & Tippet, L. (2004). Memory of myself: Autobiographical memory and identity in Alzheimer's disease. Memory, 12 (1), 56-74.

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychology , 5*, 297-305.

Armstrong, D. (1968). *A materialist theory of the mind.* London: Routledge.

Atance, C. M. (2008). Future thinking in young children. *Current Directions in Psychological Science , 17*, 295-98.

Baars, B. J. (2003). Treating consciousness as a variable: The fading taboo. In B. J. Baars, W. P. Banks, & J. B. Newman (Eds.), *Essential Sources in the Scientific Study of Consciousness* (pp. 1-10). Cambridge, MA: MIT Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Bauer, P. J. (2006). Constructing a past in infancy: A neuro-developmental account. *Trends in Cognitive Sciences , 10* (4), 175-81.

Bauer, P. J. (1997). Development of memory in early childhood. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 83-111). Sussex, UK: Psychology Press.

Bauer, P. J., & Wewerka, S. S. (1995). One- to two-year-olds' recall of events: The more expressed, the more impressed. *Journal of Experimental Child Psychology , 59*, 475-96.

Bauer, P. J., Wenner, J. A., Dropik, P. L., & Wewerka, S. S. (2000). Parameters of remembering and forgetting in the transition from infancy to early childhood. *Monographs of the Society for Research in Child Development , 65* (4).

Bennett, J. (1974). *Kant's dialectic.* Cambridge: Cambridge UP.

Blum, H. (2005). Language of affect. In S. Akhtar, & H. Blum (Eds.), *The language of emotions: Development, psychopathology, and technique* (pp. 1-18). Lanham, MD: Jason Aronson.

Bosch, G. (1970). *Infantile autism.* (D. Jordan, & I. Jordan, Trans.) New York: Springer-Verlag.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature , 391*, 756.

Boucher, J., & Bowler, D. (Eds.). (2008). *Memory in autism: Theory and evidence.* Cambridge: Cambridge University Press.

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality , 13*, 420-32.

Bowler, D. M., Gardiner, J. M., & Gaigg, S. B. (2007). Factors affecting conscious awareness in the recollective experience of adults with Asperger's syndrome. *Consciousness and Cognition , 16*, 124-43.

Broadbent, D. E., Cooper, P. F., Fitzgerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology , 21*, 1-16.

Brook, A. (1994). *Kant and the mind.* Cambridge: Cambridge UP.

Brook, A. (2001). Kant, self-awareness and self-reference. In A. Brook, & R. C. DeVidi (Eds.), *Self-reference and self-awareness* (pp. 9-30). Amsterdam: John Benjamins Pub. Co.

Brook, A., & Raymont, P. (forthcoming). *A unified theory of consciousness.* Cambridge, MA: MIT Press.

Campbell, J., Bell, S. & Keith, L. (2001). Concurrent validity of the Peabody picture vocabulary test - Third edition as an intelligence and achievement screener for low SES African American children. *Assessment, 8* (1), 85-94.

Carver, & Scheier. (1978). Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology , 36* (3), 324-32.

Ceci, S., Huffman, M. L., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about a non-event: Source misattributions among preschoolers. *Consciousness and Cognition, 3*, 388-407.

Cheyne, J. A., Carriere, J. S., & Smilek, D. (2009). Absent minds and absent agents: Attention-lapse induced alienation of agency. *Consciousness and Cognition, 18*, 481-93.

Cheyne, J. A., Carriere, J. S., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition, 15*, 578-592.

Clark, E. V. (1978). From gesture to word: On the natural history of deixis in language acquisition. In J. S. Bruner, & A. Garton (Eds.), *Human growth and development: Wolfson College lectures, 1976.* Oxford: Oxford UP.

Cohen, M., & Carr, W. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6* (4A), 383-4.

Cole, E. B., Oshima-Takane, Y., & Yaremko, R. L. (1994). Case studies of pronoun development in two hearing-impaired children: Normal, delayed or deviant? *International Journal of Language and Communication Disorders*, *29* (2), 113-29.

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

Conway, M. A., & Rubin, D. C. (1993). The structure of autobiographical memory. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 103-39). Hillsdale, NJ: Erlbaum.

Costantini, M., & Haggard, P. (2007). The rubber hand illusion: Sensitivity and reference frame for body ownership. *Consciousness and Cognition*, *16*, 229-40.

Cozby, P. (2007). *Methods in behavioral research* (9th ed.). Boston: McGraw-Hill.

Craik, F. I., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., et al. (1999). In search of the self: A positron emission tomography study. *Psychological Science*, *10*, 26-34.

Crick, F., & Koch, C. (2003). Consciousness and neuroscience. In B. J. Baars, W. P. Banks, & J. B. Newman (Eds.), *Essential Sources in the Scientific Study of Consciousness* (pp. 35-53). Cambridge, MA: MIT Press.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., et al. (1997). Looking for the agent: An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, *65*, 71-86.

Davis, D., & Brock, T. (1975). Use of first person pronouns as a function of increased objective self-awareness and prior feedback. *Journal of Experimental Social Psychology*, *11*, 381-88.

Dawson, G., & McKissick, F. C. (1984). Self-recognition in autistic children. *Journal of Autism and Developmental Disorders*, *14*, 383-94.

Dennett, D. (1981). Making sense of ourselves. *Philosophical Topics*, *12* (1), 63-81.

Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence, & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179-220). Oxford: Oxford University Press.

Dunn, L.M., & Dunn, L.M. (1997). *The Peabody picture vocabulary test - Third edition.* Circle Pines, MN: American Guidance Service.

Eacott, M. J., & Crawley, R. A. (1999). Childhood amnesia: On answering questions about very early life events. *Memory*, *7*, 279-92.

Eacott, M. J., & Crawley, R. A. (1998). The offset of childhood amnesia: Memory for events that occurred before age 3. *Journal of Experimental Psychology: General*, *127*, 22-33.

Evans, G. (1982). *The varieties of reference.* (J. McDowell, Ed.) Oxford: Oxford UP.

Exner, J. E. (1973). The self-focus sentence completion: A study of egocentricity. *Journal of Personality Assessment*, *37*, 437-55.

Fast, I., Marsden, K., Cohen, L., Heard, H., & Kruse, S. (1996). Self as subject: A formulation and an assessment strategy. *Psychiatry*, *59*, 34-47.

Ferrari, M., & Matthews, W. S. (1983). Self-recognition deficits in autism: Syndrome-specific or general developmental delay? *Journal of Autism and Developmental Disorders*, *13*, 317-24.

Ganellen, R., & Carver, C. (1985). Why does self-reference promote incidental encoding? *Journal of Experimental Social Psychology*, *21*, 284-300.

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: A first-person approach. *Philosophical Transactions of the Royal Society of London*, 1351-61.

Gardiner, J. M., Gregg, V. H., & Karayianni, I. (2006). Recognition memory and awareness: Occurrence of perceptual effects in remembering or in knowing depends on conscious resources at encoding, but not at retrieval. *Memory & Cognition*, *34* (2), 227-39.

Graham, G. (2004). Self-attribution: Thought insertion. In J. Radden (Ed.), *The Philosophy of Psychiatry: A Companion* (pp. 89-105). New York: Oxford UP.

Heydrich, L., Dieguez, S., Grunwald, T., Seeck, M., & Blanke, O. (2010). Illusory own body perceptions: Case reports and relevance for bodily self-consciousness. *Consciousness and Cognition*, *19*, 702-10.

Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review*, *104* (3), 499-523.

Howes, M., Siegel, M., & Brown, F. (1993). Early childhood memories: Accuracy and affect. *Cognition*, *47*, 95-119.

Hume, D. (1739/1978). *A treatise of human nature* (2nd ed.). (L. A. Selby-Bigge, & P. H. Nidditch, Eds.) Oxford: Oxford UP.

James, W. (1890). *The principles of psychology.* New York: Holt.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research*, *142* (1), 1-15.

Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind & Language*, *19* (2), 113–46.

Kant, I. (1781/1998). *Critique of Pure Reason.* (P. Guyer, & A. W. Wood, Trans.) Cambridge: Cambridge UP.

Karlsen, P. J., Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2010). Binding across space and time in visual working memory. *Memory & Cognition*, *38* (3), 292-303.

Kass, S. J., Beede, K. E., & Vodanovich, S. J. (2010). Self-report measures of distractibility as correlates of simulated driving performance. *Accident Analysis and Prevention*, *42*, 874-880.

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, *12*, 467-72.

Koch, C. (2004). *The quest for consciousness.* Englewood, CO: Roberts and Company Publishers.

Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1990). *The autobiographical memory interview.* Suffolk, England: Thames Valley Test Company.

Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1989). The autobiographical memory interview: A new assessment of autobiographical and personal semantic memory in amnesia patients. *Journal of Clinical and Experimental Neuropsychology*, *5*, 724-44.

Kuiper, N. A. (1982). Processing personal information about well-known others and the self: The use of efficient cognitive schemata. *Canadian Journal of Behavioural Science*, *14*, 1-12.

Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self-other differences. *Journal of Personality and Social Psychology*, *37*, 499-514.

Larson, G. E., & Merritt, C. R. (1991). Can accidents be predicted? An empirical test of the cognitive failures questionnaire. *Applied Psychology: An International Review*, *40*, 37-45.

Larson, G. E., Alderton, D. L., Neideffer, M., & Underhill, E. (1997). Further evidence of dimensionality and correlates of the cognitive failures questionnaire. *British Journal of Psychology*, *88*, 29-38.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21* (3), 451-68.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133* (3), 339-54.

Lee, A., Hobson, R. P., & Chiat, S. (1994). I, you, me, and autism: An experimental study. *Journal of Autism and Developmental Disorders*, *24* (2), 155-76.

Legrand, D. (2007). Pre-reflective self-as-subject from experiential and empirical perspectives. *Consciousness and Cognition*, *16*, 583-99.

Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, *116*, 252-82.

Lenggenhager, B., Mouthon, M., & Blanke, O. (2009). Spatial aspects of bodily self-consciousness. *Consciousness and Cognition*, *18*, 110-17.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self.* New York: Plenum.

Lewis, M., & Carmody, D. P. (2008). Self-representation and brain development. *Developmental Psychology*, *44*, 1329-34.

Lewis, M., & Ramsay, D. (2004). Development of self-recognition, personal pronoun use and pretend play during the 2nd year. *Child Development*, *75*, 1821-31.

Lind, S. E., & Bowler, D. M. (2009). Delayed self-recognition in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*, 643-50.

MacDonald, S., Uesiliana, K., & Hayne, H. (2000). Cross-cultural and gender differences in childhood amnesia. *Memory*, *8*, 365-76.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* Boston: MIT Press.

Mandler, J. M. (1990). Recall of events by preverbal children. *Annals of the New York Academy of Sciences*, *608*, 485-503.

Mandler, J. M., & McDonough, L. (1995). Long-term recall of event sequences in infancy. *Journal of Experimental Child Psychology*, *59*, 457-74.

Mathôt, S. (2010, September 25). *Online Gabor patch generator*. Retrieved September 8, 2011, from COGSCIdotNL: http://www.cogsci.nl/software/online-gabor-patch-generator

McDermott, D. (2007). Artificial intelligence and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 117-50). Cambridge: Cambridge UP.

McDonough, L., Mandler, J. M., McKee, R. D., & Squire, L. R. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Sciences*, *92 (16)*, pp. 7580-4.

Meissner, W. W. (2008). The role of language in the development of the self III. *Psychoanalytic Psychology*, *25*, 242-56.

Most, S. B., Simons, D. J., Scholl, B. J., & Chabris, C. F. (2000). Sustained inattentional blindness: The role of location in the detection of unexpected dynamic events. *Psyche*, *6* (14).

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

Nelson, K. (1993). The psychological and social origins of autobiographical memory. *Psychological Science*, *4*, 7-14.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review*, *111* (2), 486-511.

Neuman, C. J., & Hill, S. D. (1978). Self-recognition and stimulus preference in autistic children. *Developmental Psychobiology*, *11*, 571-78.

O'Neill, D. K., Astington, J. W., & Flavell, J. H. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development*, *63*, 474-90.

Oshima-Takane, Y. (1992). Analysis of pronominal errors: A case study. *Journal of Child Language*, *19*, 111-31.

Oshima-Takane, Y., & Oram, J. (1991). Acquisition of personal pronouns: What do comprehension data tell us. *International Society for the Study of Behavioral Development.* Minneapolis, MN.

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, *59*, 516-48.

Perner, J., Kloo, D., & Gornik, E. (2007). Episodic memory development: Theory of mind is part of re-experiencing experienced events. *Infant and Child Development*, *16*, 471-90.

Perry, J. (1977). Frege on demonstratives. *The Philosophical Review*, *86*, 474-97.

Perry, J. (1979). The problem of the essential indexical. In A. Brook, & R. DeVidi (Eds.), *Self-reference and self-awareness* (pp. 143-59). Amsterdam, The Netherlands: John Benjamins Pub. Co.

Peterson, C. (2002). Children's long-term memory for autobiographical events. *Developmental Review*, *22*, 370-402.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PloS ONE*, *3* (12), 1-9.

Picard, L., Ruffeveille, I., Eustache, F., & Piolino, P. (2009). Development of autonoetic autobiographical memory in school-age children: Genuine age effect or development of basic cognitive abilities? *Consciousness and Cognition*, *18*, 864-76.

Piolino, P., Desgranges, B., & Eustache, F. (2009). Episodic autobiographical memories over the course of time: Cognitive neuropsychological and neuroimaging findings. *Neuropsychologia*, *47*, 2314-29.

Piolino, P., Hisland, M., Ruffeveille, I., Matuszewski, V., Jambaqué, I., & Eustache, F. (2007). Do school-age children remember or know the personal past? *Consciousness and Cognition*, *16*, 84-101.

Povinelli, D. J. (2001). The self: Elevated in consciousness and extended in time. In C. Moore, & K. Lemmon (Eds.), *The Self in Time: Developmental Perspectives* (pp. 75-95). Mahaw, NJ: Lawrence Erlbaum Associates.

Povinelli, D. J., & Simon, B. B. (1998). Young children's understanding of briefly versus extremely delayed images of the self: Emergence of the autobiographical stance. *Developmental Psychology*, *34*, 188-94.

Povinelli, D. J., Landau, K. R., & Perilloux, H. K. (1996). Self-recognition in young children using delayed versus live feedback: Evidence of a developmental Asynchrony. *Child Development*, *67*, 1540-1554.

Povinelli, D. J., Theall, L. A., Clark, B. R., & Castille, C. M. (1999). Development of young children's understanding that the recent past is causally bound to the present. *Developmental Psychology*, *35* (6), 1426-39.

Pressley, M., & Schneider, W. (1997). *Introduction to memory development during childhood and adolescence.* Mahwah, NJ: Erlbaum.

Quoidbach, J., Hansenne, M., & Motter, C. (2008). Personality and mental time travel: A differential approach to autonoetic consciousness. *Consciousness and Cognition*, *17*, 1082-92.

Rathbone, C. J., Moulin, C. J., & A., C. M. (2009). Autobiographical memory and amnesia: Using conceptual knowledge to ground the self. *Neurocase*, *15*, 405-18.

Rochat, P., & Striano, T. (2000). Perceived self in infancy. *Infant Behavior and Development*, *23*, 513-30.

Rosenbaum, R. S., Köhler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., Black, S. E., et al. (2005). The case of K.C.: Contributions of a memory-impaired person to memory theory. *Neuropsychologia*, *43*, 989-1021.

Rubin, D. C. (2000). The distribution of early childhood memories. *Memory*, *8*, 265-9.

Rubin, D. C., & Schullkind, M. D. (1997). Distribution of important and word-cued autobiographical memories in 20-, 35-, and 70-year-old adults. *Psychology and Aging*, *12*, 524-35.

Schacter, D. L., Chiao, J. Y., & Mitchell, J. P. (2003). The seven sins of memory. Implications for self. *Annals of the New York Academy of Sciences*, *1001* (1), 226–39.

Schiff-Myers, N. (1983). From pronoun reversals to correct pronoun usage: A case study of a normally developing child. *Journal of Speech and Hearing Disorders*, *48*, 385-94.

Schneider, W., & Bjorklund, D. F. (1988). Memory. In W. Damon (Series Ed.), D. Kuhn, & R. S. Siegler (Vol. Eds.) (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5 ed., pp. 467-521). NY: Wiely.

Seager, W. (2007). A brief history of the philosophical problem of consciousness. In *The Cambridge Handbook of Consciousness* (pp. 9-33).

Shoemaker, S. (1968). Self-reference and self-awareness. In *Self-reference and self-awareness* (pp. 81-93).

Silberg, J. L. (1978). The development of pronoun usage in the psychotic child. *Journal of Autism and Childhood Schizophrenia*, *8*, 413-25.

Sinnett, S., Costa, A., & Soto-Franco, S. (2006). Manipulating inattentional blindness within and across sensory modalities. *The Quarterly Journal of Experimental Psychology*, *59* (8), 1425-42.

Spiker, D., & Ricks, M. (1984). Visual self-recognition in autistic children: Developmental relationships. *Child Development*, *55*, 214-25.

Stephens, G. L., & Graham, G. (2000). *When self-consciousness breaks; Alien voices and inserted thoughts.* Cambridge, MA: The MIT Press.

Stipek, D. J., Gralinski, J. H., & Kopp, C. B. (1990). Self-concept development in the toddler years. *Developmental Psychology*, *26*, 972-77.

Strawson, G. (2004). Against narrativity. *Ratio*, *17* (4), 428-52.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121* (3), 371-94.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). I move, therefore I am: A new theoretical framework to investigate agency and ownership. *Consciousness and Cognition*, *17*, 411-24.

Terr, L. (1988). What happens to early memories of trauma? A study of twenty children under age five at the time of documented traumatic events. *Journal of the American Academy of Children and Adult Psychiatry*, *27*, 96-104.

Tipper, S. P., & Baylis, G. C. (1987). Individual differences in selective attention: the relation of priming and interference to cognitive failure. *Personality and Individual Differences*, *8*, 667-675.

Travis, F. (2006). From I to I: Concepts of self on an object-referral/self-referral continuum. In A. P. Prescott (Ed.), *The Concept of Self n Psychology* (pp. 21-43). Hauppauge, NY: Nova Science Publishers.

Travis, F., Arenander, A., & DuBois, D. (2004). Psychological and physiological characteristics of a proposed object-referral/self-referral continuum of self-awareness. *Consciousness and Cognition*, *13*, 401-20.

Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia*, *48*, 703-12.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. S. Terrace, & J. Metcalfe (Eds.), *The Missing Link in Cognition* (pp. 4-56). New York: Oxford University Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1-25.

Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*, 385-98.

Tulving, E. (2004). How many memory systems are there? In D. A. Balota, & E. J. Marsh (Eds.), *Cognitive Psychology* (pp. 362-78). New York: Psychology Press.

Tulving, E. (1993). What is episodic memory? *Current Perspectives in Psychological Science*, *2*, 67-70.

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, *2* (3), 387-90.

Tulving, E., Markowitsch, H. J., Craik, F. I., Habib, R., & Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex*, *6*, 71-79.

Usher, J. A., & Neisser, U. (1993). Childhood amnesia and the beginnings of memory for four early life events. *Journal of Experimental Psychology: General*, *122*, 155-65.

Vandekerckhove, M. M. (2009). Memory, autonoetic consciousness and the self: Consciousness as a continuum of stages. *Self and Identity*, *8*, 4-23.

Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: Evidence against a pre-attentive subitizing mechanism. *PLoS One*, *3* (9), 1-6.

Vom Hofe, A., Mainemarre, G., & Vannier, L. (1998). Sensitivity to everyday failures and cognitive inhibition: Are they related? *European Review of Applied Psychology*, *48*, 49-55.

Weigle, T. W., & Bauer, P. J. (2000). Deaf and hearing adults' recollections of childhood. *Memory*, *8*, 293-309.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin*, *121* (3), 331-54.

Wittgenstein, L. (1934/1960). *The blue and brown books.* New York: Harper & Row.

Word Generator. (2011, November). *wordgenerator.net*. Retrieved March 30, 2012, from http://www.wordgenerator.net

Appendices

Appendix A – Acronyms

CFQ – Cognitive Failures Questionnaire

EAM – episodic-autobiographical memory

LSR – low self-reflection

HSR – high self-reflection

SAO – self-as-object

SAS – self-as-subject

SRE – self-reference effect

T/ATS – Tetris/Auditory Temporal Sequence

VA/SIB – Visual Attentiveness/Sustained Inattentional Blindness

Appendix B – Caregiver Report on Event Memory and Pronoun Use

Please answer 'yes' or 'no', and fill in the blanks where appropriate:

Has your child had a birthday party…

…in the last year but longer than six months ago?	Yes	No
…in the last six months?	Yes	No
…in the last week?	Yes	No

Has your child been to another child's birthday party…

…in the last year but longer than six months ago?	Yes	No
…in the last six months?	Yes	No
…in the last week?	Yes	No

Besides a birthday party, please specify a pleasant memorable event in general terms (e.g., trip to the zoo, concert, circus) that your child has been to…

…in the last year but longer than six months ago. _____

…in the last six months. _____

…in the last week. _____

On a scale of 0-6, with 0 being “never” and 6 being “very frequent”:

Does your child…

	Never		Occasionally				Very Frequent
…ever use general evaluative terms about himself/herself (e.g., "I'm a good girl," "Susie's pretty")?	0	1	2	3	4	5	6
…ever resist your help by saying "do it myself," "Cindy do it," or the equivalent?	0	1	2	3	4	5	6
…ever use general evaluative terms when talking about someone else (e.g., "bad dog," "Johnny's bad or mean")?	0	1	2	3	4	5	6
…ever say "I can't"?	0	1	2	3	4	5	6
…ever use descriptive terms that contain some evaluation (e.g., "sticky hands," point to toys and say "dirty" or "broken")?	0	1	2	3	4	5	6

…ever use his/her own name (e.g., "Give it to Jacob," "Jacob's truck")?	0	1	2	3	4	5	6
…ever insist on wearing certain clothing?	0	1	2	3	4	5	6
…use the word "me"?	0	1	2	3	4	5	6
…use the word "mine"?	0	1	2	3	4	5	6
…know whether he/she is a girl or boy?	0	1	2	3	4	5	6
…use the word "I"?	0	1	2	3	4	5	6
…describe himself/herself by physical characteristics (e.g., curly hair)?	0	1	2	3	4	5	6
…recognize himself/herself in the mirror (identify himself/herself by name; point to mirror when you say "where is ?")?	0	1	2	3	4	5	6
…ever call attention to something about himself/herself, like hair or clothing?	0	1	2	3	4	5	6
…communicate likes and dislikes verbally?	0	1	2	3	4	5	6
…recognize himself/herself in pictures?	0	1	2	3	4	5	6
…ever call attention to something he/she did (e.g., "Look what I did," or by gesture—showing you something she/he did)?	0	1	2	3	4	5	6

…remember events from one to two years ago?	0	1	2	3	4	5	6
…remember events from the last year to six months ago?	0	1	2	3	4	5	6
…remember events from the last six months?	0	1	2	3	4	5	6
…remember events from the last month?	0	1	2	3	4	5	6
…remember events from last week?	0	1	2	3	4	5	6

Appendix C – Picture Card Stimuli

Figure 1. *Picture card stimuli.*

Appendix D – Past-self-recognition Questionnaire

Pre-reach questions (if participant does not reach):

1. [*Pointing to image of participant*]Who is that? (‘me’/’[participant’s name]’/’you’/other:_____)

2. [*Pointing to image of sticker*]Where is that sticker right now? (‘my head’/’[participant’s name]’s head’/’his/her head’)

3. What is [experimenter’s name] doing right there? (‘putting the sticker on…’: ‘my head’/’[participant’s name]’s head’/’his/her head’)

4. *If, in response to ‘Who is that?’, the participant said:*

‘me’: [*Pointing to image of sticker*]What is that on your head?

‘[participant’s name]’: [*Pointing to image of sticker*]What is that on [participant’s name]’s head?

nothing, or uses another’s name:[*Pointing to image of sticker*]What is that on his/her head?

Post-reach questions:

1. Why did you reach for the sticker?

2. How did you know where the sticker was?

3. What made you think the sticker was on your head?

4. Who do you see in the picture?

Is it me?

Is it you?

Is it [*experimenter’s name*]?

Is it [*participant’s name*]?

5. Where were you when the picture was taken? 3

Was it here? 2

Was it in the other room? 1

Was it at home? 0

Was it at school? 0

6. When was the picture taken? 3

How long ago was the picture taken? 2

Was it just a little while ago? 1

Was it when you were in the other room? 1

Was it yesterday? 0

7. What were you doing when the picture was taken? (‘cup game’, other) 3

Were you playing? 2

Were you playing a game? 1

What game was it? 1

Appendix E – Episodic-recall Questionnaire

For picnic event:

Do you remember the picnic game?	yes / no
Who played the picnic game?	'I did' / 'you did' / 'we did' / IDK / other
Where did we play that?	incorrect / correct / IDK
When did you play that?	_____
What was the first thing you put away?	incorrect / correct / IDK
What was the second thing you put away?	incorrect / correct / IDK
What was the third thing you put away?	incorrect / correct / IDK
What was the last thing you put away?	incorrect / correct / IDK
What was your favourite thing?"	_____
What thing didn't you like?"	_____

Pre-test events:

For each event:

"Have you ever (had a birthday party / been to a birthday party / other)?"

_____

"When did it happen?"	_____
"Where did it happen?"	_____
"Who was there?"	_____
"What happened?"	_____

Appendix F – Auditory Temporal Sequence Narrative

The following is a first-person version of the narrative. The task used a second-person version and a third-person version (using a proper name instead of the third-person pronoun). All other details are identical across the different versions. Participants received the narrative in audio form.

A Trip to the Cottage

I pulled up to the cottage in my ratty old car, shut off the engine and popped the trunk. From the trunk I retrieved a large duffel bag with a leather strap. I hefted the bag onto my shoulder and walked down the short gravel path to the newly-built log cabin. The cabin was constructed from white pine and overlooked a tranquil lake, the shoreline thick with evergreens. As I walked up the steps to the porch, I noticed a beat-up cardboard box sitting outside the door. The box had a bright sticker on it, with the words “fragile” written in bold letters. Surprised, I picked up the box and held it in one arm, while I fumbled for the keys in my pocket with the other hand. I unlocked the door and stepped inside.

Inside the cabin, I dropped the duffel bag by the door and made my way to a circular oak table in the middle of the main room. A window set in the far wall let a shaft of light into the dim interior. Out the window I could see some tall mountains framed by forest. I tore a strip of packing tape from the box, breaking its seal, and looked inside. Perplexed, I lifted out a small, ornate, wrought-iron lamp with a glass shade and a plastic cord. It looked like it was intended to appear older than it actually was. As I set it down on the table beside a small rubber tree, I noticed that my finger was bleeding. I realized that the shade was partially shattered and I had cut myself on the glass.

I went into the bathroom to look for some hydrogen peroxide and a bandage. The bathroom had a black tile floor and a checkerboard pattern of white and baby blue tiles on the walls. It had a sink of modern design and a small toilet, along with an old-fashioned porcelain bathtub that looked out of place. I opened the medicine cabinet but could not find any peroxide. I did, however, find some rubbing alcohol and a cloth bandage. I also spotted a yellow bottle of shampoo that had been misplaced. After cleaning my cut and applying the bandage, I placed the shampoo on the edge of the bathtub.

My stomach rumbled, so I decided to go to the kitchen and make myself a sandwich. The kitchen was small, with only a few appliances, including a refrigerator, a toaster, an oven range, and a coffee-maker. Seeing the silver coffee-maker, I decided that I would have a cup of coffee with my sandwich. I placed a few scoops from a can of Colombian coffee grounds in the filter and filled the machine with water, then switched it on. It made a strange noise as I flipped the switch, but I thought nothing of it and began to construct my sandwich.

When I opened the door of the refrigerator, a tube of fish paste tumbled out onto the floor. I picked it up and returned it to a shelf in the fridge. The fridge had a small assortment of food – a jar of pickles, a head of lettuce, some mustard and a plastic bag containing a few slices of roast beef. I took all of these items except for the pickles and placed them on the counter. I fetched the duffel bag by the door and placed it on the table. I unzipped it and removed a loaf of bread, which I then placed on the counter. As I finished putting together the sandwich, the coffee-maker chimed to let me know the coffee had finished brewing. I noticed that the coffee had a very earthy smell. I poured myself a cup of the hot liquid and took a sip. I spat the coffee across the room – it tasted awful! I realized at that moment that what I thought were coffee grounds was actually dirt.

Appendix G – Free Recall Questionnaire

Participants were asked the following questions in random order.

Who was the main character?

Where did the story take place?

What happened when the main character opened the box?

Describe the box.

Where did the main character place the box?

Where did the main character place the duffel bag?

What kind of tree was on the table?

What was the main character looking for in the bathroom?

What did the bathroom look like?

What was misplaced in the bathroom?

What was in the refrigerator?

Describe the objects in the bathroom.

Describe the objects in the main room.

Describe the objects in the kitchen.

Describe the landscape surrounding the cabin.

Was there anything else memorable about this story?

Appendix H – Cognitive Failures Questionnaire

The following was reproduced from Broadbent *et al.* (1982). In question 23, the word "shops" was replaced by "store" to reflect the regional dialect.

The following questions are about minor mistakes which everyone makes from time to time, but some of which happen more often than others. We want to know how often these things have happened to you in the past 6 months. Please circle the appropriate number.

		Very Often	Quite often	Occasion-ally	Very rarely	Never
1.	Do you read something and find you haven't been thinking about it and must read it again?	4	3	2	1	0
2.	Do you find you forget why you went from one part of the house to the other?	4	3	2	1	0
3.	Do you fail to notice signposts on the road?	4	3	2	1	0
4.	Do you find you confuse right and left when giving directions?	4	3	2	1	0
5.	Do you bump into people?	4	3	2	1	0
6.	Do you find you forget whether	4	3	2	1	0

	you've turned off a light or a fire or locked the door?					
7.	Do you fail to listen to people's names when you are meeting them?	4	3	2	1	0
8.	Do you say something and realize afterwards that it might be taken as insulting?	4	3	2	1	0
9.	Do you fail to hear people speaking to you when you are doing something else?	4	3	2	1	0
10.	Do you lose your temper and regret it?	4	3	2	1	0
11.	Do you leave important letters unanswered for days?	4	3	2	1	0
12.	Do you find you forget which way to turn on a road you know well but rarely use?	4	3	2	1	0
13.	Do you fail to see what you want in a supermarket (although it's there)?	4	3	2	1	0
14.	Do you find yourself suddenly wondering whether you've used a	4	3	2	1	0

	word correctly?					
15.	Do you have trouble making up your mind?	4	3	2	1	0
16.	Do you find you forget appointments?	4	3	2	1	0
17.	Do you forget where you put something like a newspaper or a book?	4	3	2	1	0
18.	Do you find you accidentally throw away the thing you want and keep what you meant to throw away – as in the example of throwing away the matchbox and putting the used match in your pocket?	4	3	2	1	0
19.	Do you daydream when you ought to be listening to something?	4	3	2	1	0
20.	Do you find you forget people's names?	4	3	2	1	0
21.	Do you start doing one thing at home and get distracted into doing something else	4	3	2	1	0

	(unintentionally)?					
22.	Do you find you can't quite remember something although it's "on the tip of your tongue"?	4	3	2	1	0
23.	Do you find you forget what you came to the store to buy?	4	3	2	1	0
24.	Do you drop things?	4	3	2	1	0
25.	Do you find you can't think of anything to say?	4	3	2	1	0

Appendix I – Word Lists

All words generated online from Word Generator (2011).

Adjectives

Abrasive	Agreeable	Amused	Angry	Arrogant	Bad
Bewildered	Bored	Brave	Charming	Clever	Confident
Coordinated	Courageous	Crooked	Cute	Dependent	Discreet
Doubtful	Empty	Energetic	Fascinated	Fierce	Fortunate
Frantic	Fretful	Frightening	Funny	Gifted	Gorgeous
Guarded	Handy	Helpful	Ill-informed	Impolite	Judicious
Learned	Loud	Marvelous	Mellow	Naive	Nauseating
Outgoing	Pathetic	Possessive	Romantic	Ruthless	Silly
Spiteful	Talented	Testy	Tidy	Towering	Trite
Unruly	Useless	Utopian	Versed	Weary	Young

Dual-syllable nouns

Word set 1:

peanut	mirror	table	basket
spider	placemat	onion	tiger
painting	monkey	eagle	kitten
castle	swordfish	skateboard	werewolf
hotel	pillow	coaster	keyboard

Word set 2:

tulip	earthworm	bullfrog	donkey
windmill	desert	salmon	cigar

kettle	giraffe	acorn	zebra
airplane	lizard	bookcase	orange
window	scooter	human	sweater

Word set 3:

boulder	cellphone	firehose	ocean
insect	chicken	jacket	baby
bottle	flower	cactus	furnace
omelette	mussel	parrot	gecko
speaker	journal	noodle	trumpet

Word set 4:

kitchen	moustache	necklace	eyeball
camel	warthog	ladder	apple
rabbit	blanket	comet	liver
money	candle	thunder	marble
coffee	laundry	rocket	panther

Appendix J – Pilot Study - Episodic Recall Questionnaire

For brevity, I have only included the first set of questions for the pilot version of the episodic free recall questionnaire. The full form simply repeated the same set of questions for all four scenarios, changing the number in bold.

For the following questions, give as much detail as possible:

What was the person doing in the *first* scenario? _____

Whose face did you see in the images? _____

What did the person in the room look like? _____

What happened in the first frame? _____

What happened in the second frame? _____

What happened in the last frame? _____

List as many items in the room as you can remember:

Describe each item you saw in as much detail as possible (e.g., 'the toaster was green, had two slots, and a chrome handle'). Point form is acceptable:

Appendix K – Pilot Study – Temporal Sequence Scenarios

The following scenarios were presented in pictorial form, using pictures of the experimenter (with face replaced by either a celebrity of the participant's face) in his living room and digitally-added public domain images of various items. Each of the scenarios was designed to make sense when shown in reverse. For example, in a frame where the protagonist (*X*) is holding a box, it would be ambiguous whether he was lifting the box up or putting it down. This was meant to make it difficult for the participant to reconstruct the sequence of events if she could not remember the order of the sequence.

Event Type: Delivering/Receiving a Package

Sequence:

X places/removes item in/from box

X holds box

X puts/takes box in/from bin

Background Objects: book, window, teapot, stapler, tissue box, lamp, stereo, broom

Event Type: Using Ketchup

Sequence:

X opens/closes fridge

X takes/places ketchup from/in refrigerator

X puts ketchup on hamburger

Background Objects: plant, telephone, place-mat, pepper-grinder, clock, hair brush, chair, cushion

Event Type: Filling/Emptying a Cooler

Sequence:

X opens/closes cooler lid

X takes/places items from/in cooler

X picks up/ puts down cooler lid

Background Objects: telephone, window, pepper-grinder, stapler, tissue box, clock, cushion, broom

Event Type: Making Coffee

Sequence:

X lifts/places coffee mug from/on counter

X holds coffee mug

X fills coffee mug

Background Objects: book, plant, place-mat, teapot, lamp, hair brush, chair, stereo