The Concept Possession Hypothesis of Self-Consciousness

I: Introduction

Self-consciousness\(^1\) is being studied in children and animals using a wide variety of research paradigms but there is less than universal agreement on either what constitutes self-consciousness or on what counts as sufficient evidence of its possession by a subject. When Gallup [1970] claims to have demonstrated possession of self-awareness by chimpanzees because they recognised themselves in a mirror, is this the same concept of self-awareness that Michael Lewis [1994] refers to regarding infants’ usage of the first-person pronoun? Are organisms that demonstrate a theory of mind self-aware in the same sense as those that show the capacity for mental time travel? Addressing these types of questions is my overall project. My purpose in this paper is to suggest a framework for evaluating the research. The framework is based on a simple hypothesis that links concept possession with self-consciousness such that a demonstration of the former is sufficient evidence for the latter.

A Definition of Self-Consciousness

The notion of self-consciousness I am interested in is one in which the subject is aware of itself as a self. That is, self-consciousness in the existential sense, where the subject understands what it is to be a self and knows that it is a self. This is to be contrasted with any kind of primitive self-awareness in which an organism has access to some of its mental or physical states. Insofar as an organism may perceive (say) its own pain, it might be claimed that this is a case of self-awareness in that the organism is in a state of awareness and the content of that awareness involves the self (in this case a body part). But although there may be value in knowing whether a rabbit is aware of its own pain, this is not the type of self-awareness that I am most interested in and that I refer to in this paper. I want to know if the rabbit knows it is a rabbit. Or rather, I want to know if the rabbit knows that it is, period. What is of most interest is whether any animals have the same kind of existential self-awareness that human adults have.

\(^1\) I use the terms self-consciousness and self-awareness interchangeably in this paper
Central to my view of self-consciousness, but which I do not have room to argue for here, is the concept of agency. A self-conscious organism must have a concept of its own existence as an intentional agent. This idea already has wide currency within psychology and the philosophy of mind. For example, Bermúdez [1998] presents three categories defining psychological subjects: self-awareness of themselves as perceivers, bearers of reactive attitudes and agents. According to Lucy O'Brien [2007] “...our most basic awareness of ourselves is as performers of actions, mental and physical” [p3, my emphasis]. Philosopher William Richards [1984] defines self-consciousness as consciousness of agency, as does psychologist Gerald Stechler [1982]. The centrality of agency in self-consciousness is important for the discussions presented later.

A Yardstick for Self-Consciousness

I offer a simple hypothesis about the underlying nature of self-consciousness: there is a constitutive link between concept possession and self-consciousness. If so, then a convincing demonstration that the standard for self-consciousness has been met is a display of concept possession by the subject. This could be observed indirectly through behaviour indicating inferential reasoning or abstract thought, or indeed, any other behaviour which the researcher can convincingly argue represents concept possession. Arguing for the validity of this approach is my main task in this paper. For convenience, I will refer to this hypothesis as the Concept Possession Hypothesis of Self-Consciousness, or the CP Hypothesis for short. I will not rely on any empirical work to make my case. As my intention is to use the CP Hypothesis as a framework for evaluating the empirical research on self-awareness, basing the hypothesis on any part of that very body of work would lead to circularity. Instead I will attempt to convince the reader of the plausibility of the CP Hypothesis on purely dialectical grounds.

The basis for the CP Hypothesis is an intuition, which I believe is shared by many, that there is a strong correlation between intelligence and self-consciousness. Below I present some of the background thinking behind this intuition in order to establish the a priori plausibility of the CP Hypothesis. This will be followed up in subsequent
sections with discussions supporting the view. I want to be clear that I do not equate intelligence with self-consciousness; intelligence is about flexibility of behaviour, self-consciousness is about awareness of the self as a self, as an intentional agent. These are clearly different things. However, intuitively they appear to go hand-in-hand. Ethologists studying animal cognition do not explicitly link them and yet they tend to discuss these capabilities almost in the same breath (e.g., see Byrne, 2000, pp559-560). Indeed, the same animals we view as relatively intelligent are also those we suspect of being self-aware (chimpanzees, elephants, dolphins, etc).

Three Levels of Development

We can identify three broad layers of development on the road to self-consciousness, phylogenetically speaking: in the first level are organisms with no consciousness at all; in the second are organisms with consciousness but no self-consciousness and in the third level are organisms with self-consciousness.

Level 1 is for organisms with no consciousness. It might not be so easy to determine whether an organism has consciousness or not, but I think we can be fairly confident that there are at least some simple organisms that do not, such as single-celled organisms or bacteria. These organisms are not conscious of their surroundings and neither do they perceive them as such - except perhaps in some very loose sense, in the same way as might be said of a bear trap or robot: they 'perceive' their environment as when certain stimuli trigger pre-programmed responses. The organisms' ability to interact with the world is genetically hardwired into their control systems (brain, or nervous system or whatever) and not mediated by consciousness. Let us characterise level 1 by behaviour that is purely of a stimulus/response type.

In level 2 we have organisms that do have consciousness. Again, it's hard to know which animals to include in this category, but I tend to assume that all mammals, with their relatively advanced brains, reach at least this level. These organisms are aware of their environment – they are conscious of their surroundings. Many argue for different levels or types of consciousness (e.g. See Nelkin, 1996; or Kriegel, 2007) but for my purposes we can just allow that level 2 organisms have at least the type of consciousness that confers upon them perceptual awareness of their environment.
These organisms are capable of holding mental representations. For example, a visual image of an object is a mental representation of the object. The same could be said of other perceptual modes, such as audition; for example, the memory of the opening bars of Beethoven's Fifth Symphony would be a type of mental representation. There is a physical basis for these mental representations; they are somehow encoded in the dynamic physiology of the brain. So, associated with a mental representation is a physical representation. How the physical representation in brain physiology leads to the experience associated with a mental representation is beyond the scope of this work. I believe it is a fair assumption that some non-human animal species (at least mammals) have mental representations. Holding mental representations, such as visual images, does not imply understanding those images, as I will explore in more detail later.

In level 2 organisms flexibility of behaviour is increased compared to level 1 but is still limited – the behaviour may be more complex, but it is still underwritten by either genetic hard-wiring or by associative learning. Level 2 organisms are not intelligent in the way we routinely use that term. An example might be the stalking and attack response of a predator to the stimuli of hunger and perception of prey. It is my contention that this behaviour does not involve inferential reasoning abilities on the part of the predator. The essence of the actions of level 2 organisms is still based on a stimulus/response paradigm, though of a more sophisticated form because it is mediated by (perceptual) consciousness.

In my view being conscious does not necessarily mean being self-conscious. So now we come to the highest level in this simple taxonomy, level 3. Humans are a paradigm example of level 3 organisms. In humans we get both self-consciousness and intelligent behaviour. Humans are able to engage in the greatest level of flexible behaviour, with the ability to deliberate and solve complex problems. Humans, we

---

2 See David Chalmers' [1995] 'Facing Up to the Problems of Consciousness'
3 Some other authors, however, do appear to link the two. For example, ethologist Euan MacPhail [2000] suggests that animals such as cats and dogs are conscious in the sense that they experience feelings and a self-concept is necessary for the experience of feelings [p267]. To Dickinson & Balleine [2000] there are only two levels to consider rather than my three: “... S-R [stimulus-response] robots and cognitive creatures that are also endowed with intentional representations, affective experience, and the ability to integrate the two in consciousness...” [p201]. Philosopher Uriah Kriegel [2004] also argues that there is no consciousness without self-consciousness.
know, are also self-conscious in the relevant way – they are aware of their own existence. Let this dual capacity characterise level 3.

In this simple breakdown of developmental stages into three levels the first question might be *is there an important level missing?* More specifically, is there a level in between levels 2 and 3 in which we find *only* intelligence or *only* self-consciousness, but not both together? If that were the case then it would argue against the CP Hypothesis, which effectively says the two are co-dependent. But if there is such an intervening level, which of intelligence or self-consciousness characterises it? I think the difficulty in answering this question is one good reason to suspect that they in fact co-evolved.

**II: Concepts**

So far I have talked broadly about 'intelligence' and its connection with self-consciousness. Distilling these capacities to their essence is what leads to the CP Hypothesis as I have previously expressed it, in terms of concepts. Concept possession is necessary for inferential reasoning and abstract thought and I think it is the core element of intelligent behaviour. On the other side, self-consciousness, as I have been using the term, can be thought of as the possession of a *self-concept*, so concepts enter into both sides. Now, I take it to be relatively uncontroversial that concept possession is *necessary* for self-consciousness as, obviously, an organism must be concept-bearing in order to have a self-concept. My further claim is that concept possession is not only necessary but *sufficient* for self-consciousness. Thus, an alternative expression of the CP Hypothesis is as follows: *concept possession is necessary and sufficient for self-consciousness.* My aim in the next section is to present support for the CP Hypothesis. In preparation for this I first present a brief discussion on concepts. Of course, I cannot here present a comprehensive analysis of concepts and neither is it necessary to do so. I use the term in a straightforward way, as indicating that certain organisms are able to *understand* certain things in their world. The point is to distinguish those organisms from others that have no understanding, and can only react to the world via a stimulus/response paradigm.
Three Key Properties of Concepts

The following are fairly widely accepted properties of concepts:

1. Concepts are mental particulars [Fodor, 1998]. But there are mental particulars that are not concepts. For example, a visual image is a mental representation but it does not necessarily represent a concept.

2. Concepts do not exist in isolation: “...the grasping of a single concept requires the grasping of an entire body of concepts” [Brown, 1986].

3. Concepts imply inference, or reasoning power. According to Smith and Medin [1981] the three views of concepts (classical, probabilistic and exemplar) all acknowledge that “…concepts have the twin functions of categorization and inference” [p9, my emphasis]. That is, the ability to solve problems through reasoning (as opposed to ‘solving’ problems through associative learning).

The message to take away from point (1) is that it is possible to hold mental representations without being concept-bearing. According to Gareth Evans [1982/1994]: “The informational states which a subject acquires through perception are non-conceptual, or non-conceptualized. Judgements based upon such states necessarily involve conceptualization…” [p204]. The first sentence is pretty much what I'm asserting is the case with level 2 organisms, which do not have the capacity for abstract thought – that is, they are not concept-bearing. In my view, an organism cannot be said to be concept-bearing unless it is able to ascribe meanings\(^4\) to mental representations. In my schema, level 2 organisms cannot ascribe meanings to any of their mental representations – for example, their visual images. A mental representation that has been ascribed a meaning represents a concept grasped by the ascriber. Take as an example a predator perceiving prey. The cognition involved in this perception, if it may properly be called cognition, is on the level of affordance, as

\(^4\) Let me clarify the distinction I make between meaning and another term often used in similar contexts: content. All representations have content – that is, informational content. They can all be analysed and described in different ways. For instance a visual image might be described in terms of curves and angles, or by specifying the colour and brightness (etc.) of individual 'pixels' at a given level of granularity. But having informational content does not involve understanding. Ascribing a meaning to a representation does imply understanding – in other words, the grasping of the concept represented. A more detailed and rigorous explication of much the same point of view can be found in Dretske [1981], particularly in chapter 7: Coding and Content.
described by Gibson [1979]. On my reading of the notion of affordance, it is non-conceptual in nature\(^5\). The predator has no need for a concept of prey, it just needs to recognise that it affords the action of feeding. Again, no concept of feeding is necessary on the part of the predator in order for it to engage in the activity. The leopard might perceive some prey and will try to capture and eat it, but it need not have the concepts of PREY or FOOD. It is behaving according to a stimulus/response paradigm, albeit of a more complex form than that displayed by (say) cockroaches. Level 2 organisms are able to perceive their environment but not to conceive it, whereas level 3 organisms are. I like to say that level 2 organisms only have a perceptual field, while level 3 organisms also have a conceptual field. I will return to these ideas later.

Point (2) is fairly straightforward; it indicates that all concepts are relational – they necessarily bear relations to certain other concepts. The much stronger claim I argue in this paper is that all concepts bear a relation to the self-concept. I assert that if an organism is capable of holding any concept, it is must possess a self-concept. The self-concept is the primary concept, if you will. So, level 2 organisms have no self-concept (that is, they are not self-conscious) because they are not concept-bearing, while level 3 organisms are concept-bearing and must therefore hold the self-concept. In the next section I argue this position in more detail.

Point (3) above indicates one way the CP Hypothesis might be used to evaluate research. If concept possession is constitutive of self-consciousness, then on my account self-conscious organisms are able to reason inferentially, while organisms that are not self-conscious can't. If so then a convincing demonstration of inferential reasoning on the part of an experimental subject will constitute sufficient evidence of self-consciousness. Of course, any other incontrovertible evidence of concept possession will likewise be sufficient to prove self-consciousness.

\(^5\) “You do not have to classify and label things in order to perceive what they afford” [Gibson, 1979, p134]. Categorisation is a key element of conceptualisation, so the preceding quote implies the possibility of non-conceptual perception of affordances.
III: Three Discussions

In this section I present three discussions in support of the CP Hypothesis: an analogy with perception; an argument based on the ‘web of concepts’; and a new interpretation of Bermúdez [1998].

An Analogy with Perception

One facet of the CP Hypothesis is the impossibility of concept possession without possession of the self-concept: to be concept-bearing at all means having a concept of the self. I think perception, though far removed from conception, provides a model for this view, so I begin with a simple analogy to illustrate the point.

Anything that is perceived is perceived relative to the self; all perception is in this way self-specifying. For example, consider the visual mode of perception: whatever is seen bears a spatial relation to the perceiver. The position of the visual percept is specified in relation to the perceiver and hence the position of the perceiver is itself co-specified. Nothing can be perceived without the perceiver also co-perceiving itself. This applies not just to sight but to all of the other senses as well. In taste, touch and smell the physical body is always presented to the perceiver through direct contact with the perceptor; in hearing, like sight, a distal object is discerned as being in a location relative to the perceiver. Self-perception, of course, does not mean self-consciousness. An organism can be perceptually aware of its own physical self in just the same (non-conceptual) way as it is aware of other objects in its immediate environment. However, the self-specifying nature of perception is a good model for the likewise self-specifying nature of conception.

Just as the perceptual field always includes the perception of the perceiver, the 'conceptual field' always includes the concept of the conceivers — that is, the self. All other percepts in the perceptual field are defined in relation to the self as perceiver, and all other concepts in the conceptual field are defined in relation to the self as conceivers. Self-perception provides the frame of reference for the perceptual field. Analogously, self-conception provides the context for the conceptual field. What a
conceiver conceives is conceived relative to its own self-concept. This is one way in which I view the self-concept as indispensable to any cluster of concepts.

The Web of Concepts

I now present a case that all concepts are related to the self-concept. In this discussion we will assume we are dealing with organisms that are at least perceptually conscious, that is, they are at least conscious of their environment; they can detect affordances and can interact with their environment accordingly. I want to show that if such an organism can grasp any concept, it must have already grasped the self-concept. In essence, I am starting with at least level 2 organisms and showing how we can recognise that they are level 3 organisms. I will proceed by way of an example.

Consider the concept BLADE (imagine a caveman perceiving a sharpened piece of flint, say). If an organism has grasped the concept BLADE, that is, it understands what a blade is, then it must also have grasped the concept CUT. Cutting is what a blade does – it is impossible to grasp the concept BLADE without simultaneously grasping the concept CUT. Cutting is an action; a blade does not cut of its own accord, it must be used for that purpose by an organism. So, along with the concepts BLADE and CUT necessarily comes the concept of ACTION. Actions must be performed; there has to be a ‘doer’ or, more correctly, an agent, so the organism must also have grasped the concepts AGENCY and AGENT. Thus are we led from the simple concept BLADE through a chain of necessarily related concepts to the concept AGENT. If an organism has grasped the concept AGENT, then it must also have grasped the concept of self-as-agent. In other words it must hold the self-concept, in the sense previously defined, that is, recognition of itself as an agent. Note that I am not implying that an agent who is able to use a blade must be self-conscious; a caveman raised in a blade-wielding tribe will have blades as part of his environment and may well be able to use a blade without having any concept of a blade. And I propose no theory here of how an organism might come to acquire the concept BLADE or acquire any other concept, or indeed how it might come to be concept-bearing. My argument starts with the assumption that an organism is concept-bearing and has a particular concept (BLADE) and concludes that it must also have the concept AGENT and therefore also the self-concept.
If you are not completely comfortable with the final link, from the concept AGENT to the self-concept, consider the alternative. In that case, we would have an organism that has grasped all the previously mentioned concepts including the concept AGENT but without having grasped the concept of itself as an agent. Now, being an organism, it is a perceiver and an agent. It must therefore perceive its own agency. Of course, just perceiving its own agency does not mean it has a self-concept or any concept – an organism may be a perceiver and an agent and yet interact with its environment in a purely stimulus-response manner. But we have already assumed that the organism in our example is concept-bearing and has grasped the concept of agency. To suggest that it perceives its own agency and has grasped the concept of AGENCY but is yet unaware of itself as an agent seems like much the less plausible alternative. Presumably our organism is aware of other agents (let's again visualise a caveman and name him 'Homer'). Perhaps Homer has seen other cavemen using a blade in the action of cutting skins off animals. Homer, we have assumed, is concept-bearing and understands what he sees; he knows that these others are agents performing an action; he has grasped the concept of agency and ascribes this property to his conspecifics. As he has the same agency himself and can perceive his own agency in the same way as he perceives the agency of others, he must be able to apply the same concept to himself as easily as he applied it to others.

Of course I have only given one example and there is an obvious objection that my choice was selective in being an example of what Bolton [1977] calls a physical concept. Being a physical concept allows a physical interaction, which implies a link with the key concept of agency. However, I maintain we could start with any concept and will always find a link with the concept of agency. Bolton identifies at least three different types of concepts, physical, logico-mathematical and philosophical and gives BEAUTY as an example of a philosophical concept. So how does holding this concept necessarily imply holding the self-concept? Beauty, as the saying goes, is in the eye of the beholder – an organism has to apprehend beauty. There is no way to hold a concept of beauty without understanding that the object in question is beautiful to someone. And the apprehension of beauty is an action, which requires an agent. So, holding the concept BEAUTY implies holding the concepts AGENCY and AGENT. Of course, once again, an organism need not have a concept of BEAUTY in order to
apprehend it. For all we know, level 2 organisms might very well react
physiologically in the same way as human adults do to a vision of beauty. But I am
not talking about a non-conceptual gut reaction here – rather, I mean an understanding
of the very idea of beauty. An organism that understands the meaning of BEAUTY
will necessarily also understand a host of connected concepts including the idea of the
self. I believe a chain of connections can always be made to the concept of AGENCY,
and from there follows the link to the self-concept as argued above.

Bermúdez and the Paradox of Self-Consciousness

Many of the ideas articulated by Bermúdez [1998] in The Paradox of Self-
Consciousness are similar to those I have expressed here. Bermúdez's intention is
different to mine; whereas I am examining how it may be possible to detect self-
consciousness in animals and human infants, Bermúdez wants to resolve a paradox he
sees in self-consciousness. Nevertheless, after a brief review of the argument
presented by Bermúdez, I will show how his ideas map onto mine and how his line of
reasoning can be interpreted as support for the CP Hypothesis.

Bermúdez conceives self-consciousness in terms of mastery of the first-person
pronoun: “…self-conscious thought in the absence of linguistic mastery of the first-
person pronoun is a contradiction in terms” [p28]. The paradox for Bermúdez is based
on a perceived circularity: what comes first, mastery of the first-person pronoun or
self-conscious thought? Both seem to rely on the existence of the other. Bermúdez is
unsatisfied with any explanation of this interdependency that relies on what he calls
(after Christopher Peacocke) ‘local holism’ [p18], whereby the various abilities at the
root of self-conscious thought can only be explained in terms of each other. That, he
thinks, would violate what he calls as the Acquisition Constraint: if a given cognitive
capacity is psychologically real, then there must be an explanation of how it is
possible for an individual in the normal course of human development to acquire that
cognitive capacity. Thus, he is determined to find a way to circumvent the paradox
which does not rely on simply assuming that self-consciousness and linguistic
mastery of the first-person pronoun co-develop.
At the heart of the paradox is an element of the classical theory of content, the Conceptual-Requirements Principle: *ascriptions of content to an individual are constrained by the concepts the individual possesses*. So, if the content is an ‘I’-thought then this principle dictates that the individual must have the relevant concept relating to ‘I’ – that is, a self-concept. To break the paradox, Bermúdez rejects this principle and replaces it with a theory of non-conceptual content. Briefly described, this theory proposes the existence of *non-conceptual representational content*. The move Bermúdez makes is to suggest that the type of self-consciousness presupposed by linguistic mastery of the first-person pronoun is *non-conceptual* rather than ‘full-fledged’. He considers this non-conceptual self-consciousness as a *primitive* form of self-consciousness. Bermúdez then devotes considerable effort arguing (convincingly, I think) for the existence of non-conceptual *first-person* content (i.e. non-conceptual usage of ‘I’-thoughts) ascribable to creatures without conceptual or linguistic abilities. Successfully doing so thereby breaks the paradox of self-consciousness because the ‘primitive’ form of self-consciousness can support 'I'-thoughts (of a certain type) that do not rely on conceptual (and therefore linguistic) abilities.

Of course the foregoing is a highly condensed version of Bermúdez's very elaborate and fine-grained argument. Nevertheless, I believe I have captured the crucial elements. My task now is to show how these elements map onto the ideas I have presented in this paper. The best place to start is to compare our notions of self-consciousness. Bermúdez makes a distinction between 'full-fledged' self-consciousness and more primitive forms. It is the full-fledged variety that equates to my conception of self-consciousness. As to the so-called primitive forms as described by Bermúdez, I do not deny such things exist, but I do not wish to use the term 'self-consciousness' to describe them – this is simply because I want to keep a sharp distinction between the type of self-consciousness I'm interested in – the existential type that we humans take for granted for ourselves and wonder about in animals – and any other weaker form of awareness of self. So I view this difference between Bermúdez and myself merely as a difference in terminology, not in conception.

Bermúdez describes full-fledged self-consciousness in terms of linguistic mastery of the first-person pronoun, while for me it is concept possession that is the assumed necessary condition. According to Bermúdez, however, these are equivalent: “...there
is a constitutive link between language mastery and concept mastery” [p70]. Here I must admit to a point of departure with Bermúdez, for I do not wish to commit to the idea that concept possession depends on linguistic abilities. However, either way, Bermúdez's conception is consistent with mine in that he sees concept possession as a necessary condition for full-fledged self-consciousness and demotes what he terms non-conceptual self-consciousness to a 'primitive' form.

This non-conceptual primitive self-consciousness of Bermúdez is equivalent to the cognitive capacity I ascribed to my 'level 2' organisms. Bermúdez's theory of non-conceptual content is similar to my notion of the type of mental representations that level 2 organisms can have. That is, while they may have mental states such as visual images, they are unable to ascribe any meaning to them. Or, as Bermúdez puts it, “…states with representational content can be properly ascribed to individuals without those individuals necessarily possessing the concepts required to specify how those states represent the world” [p268]. Meanwhile, my 'level 3' organisms are those I think Bermúdez would say possess full-fledged self-consciousness.

So far I have shown a correspondence between Bermúdez's ideas and my own; including the notion that concept-possession is a necessary condition for (full-fledged) self-consciousness. As previously stated, I view this as relatively uncontroversial as it is obvious that an organism cannot have a self-concept if it is unable to hold any concepts. The further claim I make in the CP Hypothesis is that concept possession is sufficient for self-consciousness, which, although not explicit in Bermúdez, can be inferred from his arguments. Compare creatures with full-fledged self-consciousness and those with only the primitive non-conceptual kind. The only thing that separates them is concept possession. Both have access to internal representational states with self-specifying content, but only the full-fledged self-conscious creatures are concept-bearing. It follows from this that the addition of concepts to the 'primitively' self-conscious creatures is sufficient to confer on them full-fledged self-consciousness. There is no intervening transitional stage, in which organisms attain concepts but not full-fledged self-consciousness. It might be argued that there exists a transitional stage whereby the organism acquires only proto-concepts. In this case I would argue the organism has attained only proto-self-consciousness, and not the (full-fledged) self-
consciousness I have been discussing. On this reading of Bermúdez, his reasoning provides clear support for the CP Hypothesis.

IV Conclusion

The framework I propose in this paper is comprised of two components: (1) a conception of self-consciousness in which the subject understands he exists as an intentional agent, and (2) the hypothesis that concept possession is sufficient to confer self-consciousness on an individual. In an upcoming paper I use this hypothesis to settle once and for all the dispute over mirror self-recognition (MSR), by showing definitively that MSR is conclusive evidence for the existence of self-consciousness.

References


Evans, Gareth (1982/2001) Self-Identification in Brook and DiVidi, pp95-141


MacPhail, Euan (2000) *The Search for a Mental Rubicon*, in Heyes & Huber, pp253-271


