Self-Consciousness in Animals: Advantages and Problems of a Multipronged Approach

Florian Leonhard Wüstholz

Abstract

Self-consciousness in non-human animals is a complex phenomenon which raises both conceptual and methodological problems. First, what do we mean by the concept of 'self-consciousness'? Secondly, what is the best experimental approach to self-consciousness? This paper gives a short overview of the concept of self-consciousness in section 1. We can understand the concept of self-consciousness as capturing the ability of subjects to consciously think about themselves *as* themselves. If this is accurate, then it is prudent to look at a broad bundle of abilities which are related to this ability. The paper then focuses on advantages and problems of the *multipronged approach*, a kind of empirical approach which uses several different test criteria to determine whether a subject is self-conscious or not. Section 2 briefly contrasts this kind of approach with other possibilities. The main disparity is to singlepronged approaches, which have been discussed and dismissed in greater detail elsewhere [34]. Section 3 discusses the advantages of a multipronged approach, whereas section 4 is dedicated to the discussion of some of its problems.

1 What is self-consciousness?

Before we can seriously endeavour in the empirical task of determining which non-human animals (henceforth 'animals') are self-conscious, we need to be clear about the concept we use. I argue that self-consciousness is the ability of subjects to consciously think about themselves *as* themselves. I have given a more detailed argument for such an account in a recent article [34] and will not reiterate these conceptual arguments in detail. The present article constitutes a natural continuation of this recent one, which was dedicated to conceptual questions and the discussion of why singlepronged approaches are problematic, since I will here

Kriterion – Journal of Philosophy (2015) 29: 1–17.
http://www.kriterion.at

primarily discuss advantages and problems of a multipronged approach to self-consciousness, which have previously been left out.

The present concept of self-consciousness as the ability of subjects to consciously think about themselves as themselves bears some similarities to what John Locke has in mind when he describes the concept of a *person* in his *An Essay Concerning Human Understanding*:

> [W]e must consider what 'person' stands for: which, I think is a thinking intelligent being, that has reason and reflection, and *can consider itself as itself*, the same thinking thing in different times and places; which it does only by that *consciousness* which is inseparable from thinking [...].[1] [18, II.27.9., p. 208, emphasis mine]

Talking about self-consciousness as being an *ability* of conscious subjects is accurate because self-consciousness manifests itself in things that some subjects can *do* and others cannot. Especially, it is something that only *conscious* subjects can do (hence the name). So, claiming that someone is self-conscious is similar to claiming that someone has a sense of smell or a sense of humour—a specific ability of a subject. It is not like claiming that someone has a sharp kitchen knife in her drawer or a chipped tooth—something a subject possesses in her 'inventory'.

It is important to note that this use of the concept *self-consciousness* differs slightly from the one used by many cognitive scientists. This in itself is not a problem, since it merely implies that we are not necessarily talking about the same phenomenon. Let me illustrate briefly the use that I dismiss. The following claim is widespread among cognitive scientists: 'We need to distinguish between different "self-concepts" when we are talking about self-consciousness and correspondingly between different "degrees" of self-consciousness.' While I see nothing wrong in distinguishing different degrees of self-consciousness, Gennaro [13, 189–190], in line with other cognitive scientists, distinguishes degrees of self-consciousness on the purely *non-conscious* level. For instance, his notion of level 1 self-consciousness amounts to the ability of distinguishing one's own body from other physical things—what is often called 'minimal "self-consciousness"'—an ability which is possessed by many *prima facie* non-conscious beings.

I will reserve the term self-consciousness for a *conscious* ability of subjects. Therefore, self-consciousness is a sub-type of consciousness. In contrast, 'minimal self-consciousness' is merely a sub-type of thinking in a very general sense; for instance the ability to think about one's own

body and distinguish it from other things. I would go on to *cautiously* claim that my use of the term is more accurate and comes closer to the common sense meaning of 'self-consciousness' (or self-awareness).[2] When we look at the close connection between a person being self-conscious in the colloquial sense (of being preoccupied with oneself) and the more technical sense I have in mind (of being capable of having specific conscious thoughts), it becomes apparent that self-consciousness is a conscious ability. The former ability is in fact enabled by the latter, since a person without the ability to consciously think about herself as herself is incapable of worrying about her own weight, whether the shoes match the rest of the outfit, or whether being early to the meeting will make a bad impression. A subject with *only* 'minimal "self-consciousness"' is about as far from being self-conscious in the colloquial sense as I am of ever winning a Grand Slam title in tennis.

What then does it mean to have the ability to consciously think about oneself *as* oneself? Let me use some examples to illustrate the kind of ballpark we are playing in.

Alpha and the dirty mug
Alpha is sitting at the kitchen table and sees that someone has forgotten to wash their mug. She gets upset and thinks that whoever left the mug is very inconsiderate. Unbeknownst to her, it was she herself that left the mug standing around.

In this case, Alpha does think about herself but she is *not* self-conscious—despite being 'minimally self-conscious'. She does not think about herself *as* herself. Otherwise, she would have thought that she *herself* is very inconsiderate. While she has a conscious thought, she does not consciously 'pick *her*self out' in that thought. Whenever we are self-conscious, we employ a (non-descriptive) singular thought [7] and not a general thought with the existentially quantified content 'the person now having a thought is F'. The following case should illustrate the difference.

Beta and the old photo
Beta is going through the picture-book of her recently deceased grandmother. She finds a picture of a young girl smiling on her first day of school. She thinks that this girl must have been very happy when the picture was taken. Unbeknownst to her, it is a picture of herself on her first day of school.

Beta, like Alpha, thinks about herself but this time it is a conscious thought of the required (singular) type. She has a singular thought about

that girl, which happens to be herself. Unfortunately, she is quite ignorant of that fact, i.e. that it is a picture of herself, which is why it is not a case of self-consciousness. To show this, we can take into account that she would not consent to the following line of reasoning: '*That* girl was happy; therefore, *I* was happy'. And this would be required if it were a case of self-consciousness. So, what is an example of consciously thinking about oneself *as* oneself?

Gamma and the forgotten key
Gamma is standing in front of her door and is looking for her keys. She checks every pocket until she realises that the keys are still at the office. She then remembers putting them on the coffee table before taking off and forgetting to pick them up.[3]

Other familiar cases of self-consciousness are: reflecting on one's own single-handed backhand technique and how to improve it, wondering about what one would like to eat for dinner, picturing oneself in five years, reminiscing the beautiful day on the beach in Biarritz last summer, and so on. What all these things have in common is that they involve the subject consciously thinking about herself *as* herself. Thinking about the beach in Biarritz is only half as pleasurable if you are not imagining yourself being there. It also does not drastically improve the situation—rather the opposite—if you imagine the beach with someone there that might or might not be you.

There is some disagreement among contemporary philosophers on how this ability is best explained. Most prominent are accounts using a Neo-Fregean 'self-concept' [25, 26] or a non-descriptive mental file: the 'self-file' [28]. Less prominent are self-ascriptive views [4, 8, 17]. For the purposes of this paper, we do not have to delve into this debate since we are merely interested in a specific type of *ability* not in the exact nature of its mental implementation and explanation. Therefore, we can remain neutral on this question and be content with the following definition of self-consciousness:

Self-consciousness
A subject is self-conscious iff it has the ability to consciously think about itself *as* itself.

Whenever a subject shows some ability which requires it to consciously think about itself as itself, there is good reason to suppose that it is self-conscious. For instance, Gamma cannot remember that *she* left *her* keys on the coffee table without being able to think about herself as herself.[4]

2 Two types of approaches

Generally, we can distinguish between two types of experimental approaches to self-consciousness. On the one hand, we can use a single test to decide whether a given individual is self-conscious. This is the *singlepronged* approach. On the other hand, we can opt for a combination of tests in order to determine whether a subject is self-conscious or not. This is the *multipronged* approach.

First, let us look at three possibilities for singlepronged approaches. One philosophically prominent approach is the linguistic one: If a subject can correctly use the first-person pronoun, it is self-conscious. The general idea is that our public language mirrors our thinking and the other way round. Hence, language use is a good tool to indicate mental abilities. Such an approach need not claim that language is *necessary* for thinking. Instead, it can be understood as a mere methodological tool.[5]

Another possible approach is the neurobiological one. Just as we can find the part of the brain that is responsible for episodic memory, we should be able to identify the neural correlates of self-consciousness [5, 23].[6] The move is then to claim that any subject with a homologous brain area is *ipso facto* self-conscious.

The third singlepronged approach uses a non-linguistic form of behaviour as an indicator: for instance the mirror test [10]. In this test, a subject is accustomed to mirrors, then secretly marked with a coloured dot on its forehead, and then presented with a mirror again. If it tries to remove the dot from its own body, it passes the mirror test and if it does not, it fails the test.

All three approaches use a single criterion to test for the presence of self-consciousness in a subject. There is hence only one pass–fail question that needs to be empirically investigated. As such, there is no need to theoretically reconcile several related abilities. These approaches therefore avoid the much dreaded question of explanatory parsimony (see section 4).

Now, there is also the possibility of combining various approaches into a 'test battery'. The idea behind this multipronged approach is that self-consciousness is an ability which enables subjects to do a whole bunch of things. By testing their aptitude for performing certain actions, e.g. planning one's own actions, mastering forms of social cognition, recognising oneself in the mirror, having higher-order mental states about others' mental states, and so on, we can assess whether they are self-conscious. The general methodological move is abductive and unificatory: while we might explain the subject's aptitude for performing these actions in a

'non-self-conscious' way, the explanation which attributes self-consciousness to the subject is easier and theoretically less demanding.[7]

The multipronged approach has three steps. First, it argues that a group of abilities $\phi_1, \phi_2, \phi_3, \ldots, \phi_n$ are related to the ability to consciously think about oneself as oneself. In this specific case, the approach argues that the abilities in question have a common cause in the subject's ability to consciously think about herself as herself. Since self-consciousness, i.e. a subject's ability to consciously think about herself as herself, cannot be observed directly, other relevant observable abilities need to be linked to self-consciousness. This is the conceptual and theoretical step. Secondly, it shows that individuals of a given species have these abilities (or a significant portion of it). This constitutes the experimental step. Thirdly, it argues that the best explanation of this performance—given that the abilities are related to self-consciousness—is that individuals of this species are self-conscious. This is the explanatory step.

3 Some advantages of the multipronged approach

There are at least three advantages of the multipronged approach. The first advantage concerns the fact that many different abilities are expressions of self-consciousness. Since self-consciousness cannot be witnessed directly, we require some other observable abilities which have been linked on theoretical grounds to the ability to consciously think about oneself as oneself. The claim is *not* that each of these abilites is sufficient or necessary for self-consciousness. Some or all abilities might be individually present *without* the subject being self-conscious. Instead, the claim is that self-consciousness typically manifests itself in these abilities. I will illustrate this point with an incomplete list of different abilities that are related to self-consciousness.

- *Planning* our own future actions requires at least some understanding of ourselves. We need to be able to keep track of our own goals and our past actions. For instance, if we want to cook a nice seitan burger for dinner, we need to go to the shop first. While in the shop, we need to think about our future actions and buy the required ingredients. At home, we need to remember our initial plan in order to know what the bought ingredients are for.

- *Social cognition* is a very broad ability. One part of it is keeping track of other individuals of one's social group and their relation to oneself. For instance, male cichlids (*Astatotilapia burtoni*), a

species of fish, are capable of inferring their own relative strength from observing the outcomes of others' fights, without actually fighting all rivals [15, 33, p. 38–39].

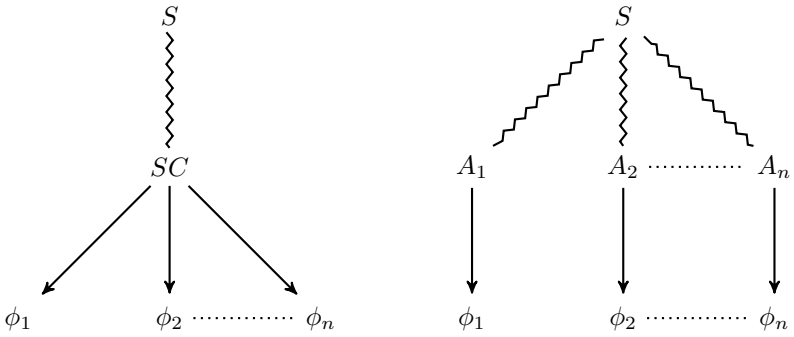
- *Mirror self-recognition* has already been described as intimately connected to self-consciousness in section 2. The ability of recognising oneself in the mirror is plausibly taken as a prime example of the expression of self-consciousness. It is unsurprising then that it is the standard behavioural test for self-consciousness.

- *Mindreading* is the ability to have higher-order mental states about others' mental states. If I see Epsilon trying to grab a piece of food, I might think *Epsilon wants that piece of food.* It is plausible to claim that an understanding of mental state predicates like *wanting*, *believing*, *seeing*, etc. is closely connected to being capable of consciously appreciating these mental states in oneself—to consciously think about oneself as oneself.

- *Using the first-person pronoun* is certainly a good indicator of self-consciousness. However, the move is not straightforward. Sign-using gorillas [24] or chimpanzees [11, 12] might not understand the signs they use. Furthermore, the move from sign-using to consciousness is generally problematic [30].

If an animal exhibits all these abilities, a proponent of the multipronged approach can give a simple explanation of the animal's capabilities: it is self-conscious. She does not need to give a different explanation for every individual ability. Instead, she can directly attribute self-consciousness to the subject which explains the various capabilities (see figure 1 on page 8).[8] Imagine a given subject, Zeta, being capable of planning, social cognition and using the first-person pronoun. We could explain why Zeta has these abilities by attributing self-consciousness to her. The *reason* why Zeta can plan her actions is that she is capable of thinking about herself *as* herself. If we did not use this multipronged approach, every single ability would require an individual, possibly independent, explanation—resulting in a much more complex theory.

The second advantage consists in the fact that the multipronged approach does not aim at providing necessary *and* sufficient conditions for self-consciousness. Instead, the argument is much more subtle. Given a range of relevant abilities, self-consciousness *best explains* the overall picture. It does not need to claim that every single relevant ability needs to be present in an animal and it also does not need to hold that every animal that exhibits all the relevant abilities *ipso facto* is self-conscious.



(1) Unification (2) Reduction

Figure 1: Explanation (arrows) of different abilities $\phi_1, \phi_2, \ldots, \phi_n$ according to the unification strategy (1) and the reductionist approach (2). The unification strategy of the multipronged approach explains the abilities by attributing (zigzag line) *one* ability (i.e. self-consciousness SC) to a subject S as a common cause for $\phi_1, \phi_2, \ldots, \phi_n$. The reductionist approach explains the same abilities by attributing (zigzag lines) *distinct* (non-self-conscious) abilities $A_1, A_2, \ldots A_n$ to a subject S as *distinct* causes for $\phi_1, \phi_2, \ldots, \phi_n$.

For instance, we might observe that a given species is capable of social cognition, planning and mirror self-recognition, but *not* mindreading. We could either explain these abilities by attributing self-consciousness to the individuals as a common cause or we could give an independent explanation for every distinct ability. What kinds of reasons speak in favour of the first route? We might argue that the presence of these three abilities is sufficient for attributing self-consciousness because such an explanation would be more theoretical parsimonious and unified—a brand of Ockham's razor. We could also observe that independent explanations of these three abilities would require other abilities of the subject which there is little empirical evidence for, interfere with each other, or are theoretically cumbersome. For instance, it is possible to explain the success in the mirror test by some elaborate non-self-conscious ability. However, such an explanation might be very implausible; e.g. it could

require a very sophisticated learning ability for which there was less empirical evidence than for the more 'demanding' ability to consciously think of oneself as oneself.[9]

The case above also shows that not every relevant ability is a necessary condition for self-consciousness. Since we might have good theoretical reasons to favour a unified explanation which attributes self-consciousness to the subject without the ability of mindreading being present, we would conclude that mindreading is *often* present in self-conscious subjects, but not always. In contrast, we might come across a subject which has all the relevant abilities. Nonetheless, we might be reluctant to attribute self-consciousness to it—for instance because it is 'just' a sophisticated and presumably unconscious robot. This shows that the multipronged approach is much more subtle and judicious than the singleprongled approaches, which use a fragile pass–fail paradigm.

The third advantage of the multipronged approach is the fact that it does not succumb as easily to the threat of reduction. It is rather easy to reduce a single behavioural ability to a non-self-conscious ability (as demonstrated in the case of the mirror test [34] or in the case of mindreading [27]).[10] However, due to the holistic nature of the multipronged approach, reduction is much more difficult. The more complex and extensive the different related abilities are, the more individual reductions of each and every one of these need to be provided. And even if such a complex and global reduction were achieved, we might still end up with a much less parsimonious explanation of the abilities of a specific species.

Let me recapitulate the three advantages of the multipronged approach over its singlepronged competitors. First, it makes use of a broad range of abilities and gives a detailed and encompassing picture of the animals in question. Secondly, it avoids a problem that is lethal for singlepronged approaches: it does not search for *one* necessary *and* sufficient condition for self-consciousness. Instead, the approach is more subtle and less prone to counterexamples. Thirdly, it is more resistant to reductive explanations. Reducing a single ability—e.g. mirror self-recognition—might be quite simple [34]. However, reducing all relevant abilities from the 'test battery' could result in a much more complex theory than unification within one overarching ability: self-consciousness.

4 Some problems and possible solutions

Despite these advantages, there are at least two important problems for the multipronged approach. First, the argument from parsimony—i.e.

the third advantage—is not always as clear as its proponents want to sell it. It is very difficult to objectively determine which of two competing explanatory theories is more parsimonious. Secondly, the so-called 'Morgan's canon' [21] gives us incentive to prefer lower-level explanations over higher-level explanations if they are available. Reduction, when possible, should always be preferred.

Let us start with the problem of parsimony. How should we decide which of two theories that predict and explain the data equally well is to be preferred? There are different kinds of parsimony which we might invoke as a standard. We might prefer simplicity, elegancy, unity, productivity, or how basic a theory is (among *many* other things).[11] *Prima facie*, there is no reason to favour e.g. a more unified over a more complex theory. So, how can we make progress? Elliott Sober has provided important insight into these epistemological problems [31] and her general arguments will here be applied to the specific case of self-consciousness.

Imagine we want to compare a simple unified model M, e.g. money explains everything, with a more complex model $M*$, e.g. money, karma, weather, and the amount of beer in my glass explains everything. Let M and $M*$ have roughly the same explanatory and predictive power. Which of the two should we prefer? The simple answer is: M, because it is simpler. Information theory confirms this initial hunch. It tells us that we need to take into account the number of adjustable parameters in a model as well as its predictive accuracy. Since $M*$ has more adjustable parameters, it gets a bigger 'penalty' than the simpler M. So, if M and $M*$ have the same predictive accuracy, we should favour M.

If, however, there is a significant difference in their explanatory power, we need to estimate the *unbiased* predictive accuracy $E(M)$ of a model. For this, we can use the (simplified) formula

$$E(M) = log\{Pr[D|L(M)]\} - k, \tag{1}$$

where k is the number of adjustable parameters in M, D is our measured data, and $L(M)$ is a specific instantiation of M where the free parameters have been assigned specific values. $Pr[D|L(M)]$ gives us the probability of observing the data given our specific model. So, $E(M)$ reflects the overall value of the model and takes into account both its predictive accuracy and its complexity. The higher $E(M)$, the better our model.[12] If M and $M*$ differ in predictive accuracy, we can use (1) as a way to objectively compare the two models.

Unfortunately, there is of yet no straightforward application of this method directly to the problem of self-consciousness in animals. However,

an example might illustrate its potential. Let M_{SC} be our unified self-consciousness model and M_R the complex reductive model represented in figure 1 on page 8. Let us assume that M_{SC} has only one adjustable parameter, whereas M_R has four. Furthermore, let the probability of observing the data given M_{SC} be 0.65 and that of M_R be 0.97. So, our reductive model predicts the data more faithfully. Nonetheless, if we calculate the *unbiased* predictive accuracy using the simplified formula (1), we get a value of -1.1871 for $E(M_{SC})$ and -4.0132 for $E(M_R)$. Accordingly, information theory would tell us that we should choose M_{SC} over M_R.

The second problem for the multipronged approach concerns the well-known principle of comparative psychology called *Morgan's canon*:

> In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale. [21, p. 53]

Unfortunately, the multipronged approach is guilty as charged. Since it does *not* claim that all relevant abilities can *only* be explained by self-consciousness, there is a 'lower-level' explanation for each of them. Planning, social cognition, mirror self-recognition, mindreading, and so on can all be explained without ascribing self-consciousness. But, how are we justified to invoke self-consciousness as an explanation if these abilities can be explained by lower-level abilities? If there are reasons to refrain *in principle* from using a 'higher-level' model where a more complex 'lower-level' model is available, the value $E(M)$ of each model should be irrelevant.

The example above gave us reason to favour M_{SC} since its unbiased predictive accuracy is higher. However, all the data can be explained by M_R which only makes use of lower-level abilities. Accordingly, Morgan's canon tells us that we should choose M_R, regardless of its unbiased predictive accuracy. Thus, Morgan's canon is a general threat to any unified higher-order explanatory endeavour. Self-consciousness might be a nice and handy way to explain a variety of related abilities, but in the end, they can all be explained without invoking the ability to consciously think about oneself as oneself. This sounds like grim news for the multipronged approach. However, as has been demonstrated by Elliott Sober [31], Morgan is aware of this problem and adds in the second edition of his *Introduction to Comparative Psychology*:

> To this, however, it should be added, lest the range of the

> principle be misunderstood, that the canon by no means excludes the interpretation of a particular activity in terms of the higher processes, if we already have independent evidence of the occurrence of these higher processes in the animal under observation. [22, p. 59]

This suggests that we should not give up the theoretical virtue of comparing complex reductive with simple unified models, for instance by employing information theory or some other tool. Morgan's canon—at least when considering the expanded version above—does not commit us to reduce reducible 'higher-level' explanations at any cost. While the dogmatic first version might undermine any kind of unificatory explanation, the same cannot be said about the more reasonable second version which allows for higher-order explanation in justified cases.

5 Many roads to self-consciousness

I argued in section 1 that self-consciousness is best understood as the ability to consciously think about oneself as oneself. I distinguished this use from a broader use common in cognitive science according to which there is self-consciousness even at the most basic non-conscious level. In section 2, I then distinguished two types of approaches to the experimental question whether some animals are self-conscious. The rest of the paper was dedicated to the multipronged approach which makes use of an array of abilities relevant for self-consciousness. I identified the advantages of such an approach in section 3: it is broad in scope, it does not search for *one* necessary *and* sufficient condition, and it is more resistant to reduction. Section 4 then discussed two problems of such an approach and how they could be overcome. By using the many roads to self-consciousness in the multipronged approach, we get an accurate and encompassing picture of self-consciousness. And only such a picture will ultimately be useful in answering the complex question concerning which animals are self-conscious.[13]

Notes

1 I do not, however, go as far as Locke in his empiricist claim that thinking is inseparable from consciousness of thinking. Furthermore, I want to leave the question of what constitutes a *person* open and do not claim that any self-conscious subject is a person or *vice versa*.

2 I do not claim that the reference of the more minimal concept of 'self-consciousness' in use by Gennaro is empty. The ability he describes is certainly present

in many subjects and most likely plays an important role in what I call 'self-consciousness'.

3 There is an interesting difference between two ways of linguistically reporting Gamma's remembrance. Gamma can remember *putting* the keys on the coffee table or she can remember *that* she put them on the coffee table. Only the former is an unambiguous report of her first-personal experience and a true case of self-consciousness. The latter could also report something she has learned by watching herself on a video surveillance tape. A classic analysis of this difference is by Castañeda [3] and the phenomenon is especially prevalent in the semantics of PRO construction [32, Chapter 3].

4 This is an instance of what Castañeda calls a quasi-indicator, which is especially important in self-consciousness as discussed in endnote 3.

5 There is a more serious philosophical position which we might call *lingualism*, that claims that only language-using subjects can entertain thoughts at all. The most popular lingualist argument is by Donald Davidson and claims (among other things) that animals are incapable of having beliefs, since they lack the holistic conceptual network required for entertaining a specific belief like *the cat went up the tree* [6]. Glock provides one of the many refutations of these kinds of argument [14]. Bennett & Hacker are also defenders of the view that language is necessary for self-consciousness [2, p. 334]. However, a proponent of the linguistic approach need not subscribe to lingualism, since she can hold on to the linguistic approach only for methodological reasons.

6 This is technically not accurate. What is searched for is usually a neural correlate of self-*representation*. However, the ability to consciously think about oneself as oneself has *prima facie* nothing to do with representing oneself. This goes somewhat against what I have argued for in the conceptual part of [34]. However, Andreas Kemmerling has argued that whatever representation we come up with as a candidate for self-representation, it can never have the same function as self-consciousness [16]. Our ability to consciously think about ourselves as ourselves is much more basic than the ability to form self-representations. In general, it seems possible to think consciously about oneself as oneself without *ipso facto* representing oneself. The case of Gamma from section 1 could serve as an example of this. If we accept Kemmerling's arguments, then the presence of a neural correlate for self-representation would be neither a necessary nor a sufficient condition for self-consciousness. The aforementioned self-ascriptive views of self-consciousness also jettison the notion of a self-representation.

7 To be more precise, the methodological move is that of a common cause abduction [29]. Self-consciousness, i.e. the ability to consciously think about oneself as oneself, is supposed to be the common cause of other dispositions and abilities. For instance, the ability to think about oneself as oneself enables a subject to pass the mirror test or correctly use the first-person pronoun.

8 The example of the mirror test helps to illustrate the nature of the distinction between the unification and reduction strategy in figure 1 on page 8. The mirror test tells us that some subjects have acquired the ability to correctly interact with their mirror images. So, when they see a spot on their reflection's forehead, they try to remove it from their own body and not from the one in the mirror. Proponents of the unification approach could then attribute self-consciousness to the subject in order to explain this ability (citing other abilities of the subject which are *also* explained by attributing self-consciousness). In contrast, the

reductionist approach would argue that this ability is explained by some other non-self-conscious ability which we need to attribute to the subject. For instance, one could argue, as I illustrated in [34], that the subject has merely acquired the ability to (non-self-consciously) interact with their own bodies in certain unusual perceptual situations in just the same way as they (non-self-consciously) interact with their own bodies in more normal perceptual situations.

9 I would like to thank Pascale Anna Lötscher for pressing me on this point.

10 The claim is not that these reductions are successful or convincing. The claim is rather that the reductions are possible and the decisions whether they are also successful are made on other grounds [9].

11 In fact, it can be argued that these kinds of epistemic scientific standards of parsimony are not purely epistemic at all. Rather, they are shaped by political and social values. For instance, feminist epistemologists such as Helen Longino argue that we should connect these 'epistemic' values with feminist values [19, 20]. This critique of traditional scientific standards of parsimony is favourable to the general argument in this section. I would like to thank an anonymous reviewer for this suggestion.

12 This is a very simplified adaption from Hirotogu Akaike [1] taken over from Elliott Sober [31, p. 243–244].

13 This paper is a very condensed version of a paper I presented under the title 'Self-Consciousness in Animals: Problems and Advantages of Various Approaches' at the Summer School *Animal Minds* in Häusern, Germany in 2014. I would like to thank the participants for helpful comments, especially Markus Wild and Hans-Johann Glock. Many thanks also go to two anonymous reviewers for useful comments and Pascale Anna Lötscher for her valuable help.

Florian Leonhard Wüstholz
University of Fribourg

<florian.wuestholz@unifr.ch>
<https://unifr.academia.edu/wuestholz>

References

[1] Hirotogu Akaike. Information Theory and an Extension of the Maximum Likelihood Principle. In Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa, editors, *Selected Papers of Hirotugu Akaike*, Springer Series in Statistics, pages 199–213. Springer New York, 1998.

[2] Max R. Bennett and Peter M.S. Hacker. *Philosophical Foundations of Neuroscience*. Wiley-Blackwell, 2003.

[3] Hector-Neri Castañeda. 'He': A Study in the Logic of Self-Consciousness. *Ratio*, 8:130–157, 1966.

[4] Roderick Chisholm. *The First Person: An Essay on Reference and Intentionality*. University of Minnesota Press, 1981.

[5] Patricia S. Churchland. Self-Representation in Nervous Systems. *Science*, 296(5566):308–310, 2002.

[6] Donald Davidson. Rational Animals. *Dialectica*, 36(4):317–328, 1982.

[7] Gareth Evans. *The Varieties of Reference*. Oxford University Press, 1982.

[8] Neil Feit. *Belief About the Self: A Defense of the Property Theory of Content*. Oxford University Press, 2008.

[9] Simon Fitzpatrick. The primate mindreading controversy: a case study in simplicity and methodology in animal psychology. In Robert W. Lurz, editor, *The Philosophy of Animal Minds*, pages 258–277. Cambridge University Press, 2009.

[10] G.G. Gallup. Chimpanzees: self-recognition. *Science*, 167(3914):86–87, 1970.

[11] Allen R. Gardner and Beatrix T. Gardner. Teaching Sign Language to a Chimpanzee. *Science*, 165(3894):664–672, 1969.

[12] Beatrix T. Gardner and Allen R. Gardner. Evidence for sentence constituents in the early utterances of child and chimpanzee. *Journal of Experimental Psychology General*, 104(3):244–267, 1975.

[13] Rocco J. Gennaro. Animals, Consciousness, and I-Thoughts. In Robert W. Lurz, editor, *The Philosophy of Animal Minds*, pages 184–200. Cambridge University Press, 2009.

[14] Hans-Johann Glock. Can Animals Judge? *Dialectica*, 64(1):11–33, 2010.

[15] Logan Grosenick, Tricia S. Clement, and Russell D. Fernald. Fish can infer social rank by observation alone. *Nature*, 445(7126):429–432, 01 2007.

[16] Andreas Kemmerling. Selbstbewusstsein ohne Selbstrepräsentation. In Hans Jörg Sandkühler, editor, *Selbstrepräsentation in Natur und Kultur*, pages 21–36. Peter Lang, Bern, 2000.

[17] David Lewis. Attitudes *De Dicto* and *De Se*. *Philosophical Review*, 88(4):513–543, 1979.

[18] John Locke. *An Essay Concerning Human Understanding*. Oxford World's Classics. Oxford University Press, 2008. Abridged with an Introduction and Notes by Pauline Phemister.

[19] Helen Longino. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press, 1990.

[20] Helen Longino. Cognitive and non-cognitive values in science: Rethinking the dichotomy. In L.H. Nelson and J. Nelson, editors, *Feminism, Science and the Philosophy of Science*, pages 39–58. Kluwer, 1996.

[21] C. Lloyd Morgan. *An Introduction to Comparative Psychology*. London: Walter Scott Publishing, 1st edition, 1894.

[22] C. Lloyd Morgan. *An Introduction to Comparative Psychology*. London: Walter Scott Publishing, 2nd edition, 1903.

[23] A. Newen and K. Vogeley. Self-representation: Searching for a neural signature of self-consciousness. *Consciousness and Cognition*, 12(4):529–543, 2003.

[24] Francine Patterson and Eugene Linden. *The Education of Koko*. New York: Holt, 1981.

[25] Christopher Peacocke. *Truly Understood*. Oxford University Press, 2008.

[26] Christopher Peacocke. *The Mirror of the World*. Context and Content. Oxford University Press, 2014.

[27] Derek Penn and Daniel J. Povinelli. On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society*, 362:731–744, 2007.

[28] François Récanati. *Mental Files*. Oxford University Press, 2012.

[29] Gerhard Schurz. Patterns of abduction. *Synthese*, 164(2):201–234, 2008.

[30] John R. Searle. Minds, brains and programs. *Behavioural and Brain Sciences*, 3(3):417–457, 1980.

[31] Elliott Sober. Parsimony and Models of Animal Minds. In Robert W. Lurz, editor, *Philosophy of Animal Minds*, pages 237–257. Cambridge University Press, 2009.

[32] Jason Stanley. *Know How*. Oxford University Press, 2011.

[33] Markus Wild. *Fische: Kognition, Bewusstsein und Schmerz. Eine philosophische Perspektive*. Bundesamt für Bauten und Logistik BBL, 2012.

[34] Florian L. Wüstholz. Selbstbewusstsein bei Tieren: begriffliche und methodologische Probleme. *Studia Philosophica*, 72:87–101, 2013.