



The expressive case for animal self-consciousness

Ángel García Rodríguez¹

Accepted: 15 February 2023
© The Author(s) 2023

Abstract

An obstacle for the attribution of self-consciousness to animals is that they lack the linguistic ability to use the first-person pronoun. To overcome the obstacle, current tests rely on the availability of behavioural measures of self-consciousness in the absence of language. However, this is not sufficient, for unless a distinction is drawn between epistemic and expressive varieties of self-consciousness, further puzzles threaten the validity of the research. This paper defends the distinction and shows how to re-evaluate current research into animal self-consciousness in its light, so as to improve conceptual clarity in this area.

Keywords Episodic memory · First-person pronoun · Mirror self-recognition · Uncertainty monitoring

1 Introduction

Self-consciousness can be thought of as the ability to distinguish between self and non-self, in a broad sense that includes the separation from one's surroundings as well as the awareness of oneself as the subject of one's own mental states; or equivalently for current purposes, the ability to be aware of one's mental states as one's own.¹ Children acquire the ability progressively through the normal process of ontogenetic development; so, cognitive deficits and pathologies aside, human adults are self-conscious. But what about non-human animals (henceforth, animals): are they self-conscious, at least some species? In asking this question, it is assumed that,

¹ It is sometimes suggested that a distinction should be drawn between consciousness of oneself and consciousness of one's mental states (Cassam, 1994). In addition, self-consciousness is sometimes understood as the ability to be aware of oneself non-accidentally, so as to accommodate such examples as Perry's shopper (1979). Both suggestions are glossed over, as they do not feature in the literature on animal self-consciousness that is the focus here.

✉ Ángel García Rodríguez
agarcia@um.es

¹ University of Murcia, Murcia, Spain

as is generally agreed, animals have mental states, whatever their exact nature and content. Therefore, only self-consciousness is the issue here.

Whether affirmative or not, the answer to the question is dependent on how animal self-consciousness is conceived. A strand of the recent debate between sceptics and non-sceptics about the attribution of self-consciousness to animals helps make the point. According to one prominent sceptical answer (Heyes, 1994), passing the mirror self-recognition test, an early paradigm for empirical research into this matter, does not require self-consciousness; rather, it is sufficient for the animal to “distinguish extraneous visual input from visual and kinesthetic feedback from its body” (915). The thought here is that, whatever self-consciousness is, it is not awareness of one’s body, and as the latter is what the test measures, passing it provides no evidence of self-consciousness.

Some anti-sceptical replies have taken the point on board and used it to argue in favour of a “gradualist” (de Waal, 2019) conception of animal self-consciousness. Thus, Bekoff (2003) has drawn a distinction between the rich (i.e. introspective) human sense of self and the parsimonious animal sense of self, the latter based on awareness of their own body whilst navigating the world. Likewise, Bekoff and Sherman (2004) have argued for a continuum of degrees of self-cognizance, ranging from the matching of phenotypic characteristics (say, odours) in oneself and others, to the bodily awareness of oneself vis-à-vis others, and the awareness of one’s own mental states. To a similar overall effect, DeGrazia (2009) has talked about types of self-awareness, from the bodily awareness typical of proprioception, sensation and agency, to the awareness of one’s social standing, and the introspective awareness of one’s mental states. Finally, Birch et al. (2020) have claimed that the division of self and non-self admits of degrees, from the separation between internal (bodily) and external events, to the conception of one’s body as temporally extended, and the conception of oneself as a persisting subject of mental states. On the whole, what the gradualist conception aims to do is make available a nuanced set of conceptual tools to assist empirical research into the question of animal self-consciousness, without setting the bar so high that a negative answer is a foregone conclusion.

This paper shares in the aspiration to improve conceptual clarity in this area. To this effect, a hitherto neglected distinction between two varieties of self-consciousness is defended – namely, the ability to reflect upon vs. the ability to express (non-rationally) one’s mental states. The neglect lies in the fact that contemporary empirical research has focused on the former variety alone; whereas both varieties of self-consciousness should be considered. Importantly, as the distinction applies to bodily and mental properties, the results of this paper can be read either as an amendment to gradualism (if the latter is thought of in reflective terms) or as a clarification of its basis.

Turning to some of the details, the paper is structured as follows. To begin with, a prominent obstacle standing in the way of the attribution of self-consciousness to animals is brought to the fore – namely, that they lack the linguistic ability to use the first-person pronoun, typically associated with self-consciousness. The focus here is on the strategy followed in recent times to overcome the obstacle – that is, the search for functional analogues of the use of the first-person pronoun. The starting point of the paper is that the strategy gives rise to two puzzles, which are a hindrance

in our quest for clarity (Section 2). So, the aim is to solve the puzzles; but how? According to the intuition to be spelled out in the rest of the paper, the source of the puzzles is conceptual; hence, it is the very idea of animal self-consciousness that needs to be clarified. It is for this purpose that the distinction between two varieties of self-consciousness, one epistemic and the other expressive, is introduced (Section 3). Subsequently, the distinction is applied to current empirical research into animal self-consciousness, by distinguishing the question whether animals are epistemically self-conscious (and the evidence for it) from the question whether they are expressively self-conscious (and the evidence thereof), particularly in the three main experimental paradigms in use nowadays (Section 4). Upon this basis, it is concluded that the puzzles hindering our clarity in this area arise only if the expressive variety of self-consciousness is overlooked; whereas recognizing that both varieties are at stake is sufficient to neutralize them (Section 5). The overall result is the expressive case for the attribution of self-consciousness to animals announced in the title.

2 One obstacle and two puzzles

The fact that animals lack the linguistic ability to use the first-person pronoun competently is commonly thought of as an obstacle for attributing self-consciousness to them. The obstacle can be construed in two ways.² On one construal, the linguistic ability is a necessary requirement for the possession of a self-concept, itself a requirement for self-consciousness. As a result, animals fail to be self-conscious. This conclusion rests on the necessary link between having a self-concept and the use of the first-person pronoun; but so conceived, the link is too strong. For, apart from the first-person pronoun, in the right circumstances other linguistic and non-linguistic resources manifest possession of a self-concept; for instance, answering questions about oneself by a nod or a shake of the head, as well as the appropriate use of definite descriptions (e.g. “the current speaker”) or one’s name.³ To be sure, such resources could be argued to depend on the use of the first-person pronoun, if they are to manifest possession of a self-concept; at least in adult humans. But this need not be pursued any further, as a weaker version of the link between having a self-concept and the use of the first-person pronoun is sufficient for present purposes.⁴

Thus, according to another construal of the linguistic obstacle under consideration, though possession of a self-concept is necessary for self-consciousness, the use

² Not all uses of the first-person pronoun count as competent; e.g. cases of mere parroting and incipient uses of the first-person pronoun by young children. Therefore, the reference to competent use is pertinent and must be read as given hereafter.

³ The qualification “in the right circumstances” is important, because not all alternatives to the first-person pronoun indicate self-consciousness. This includes young children’s use of their own name; for arguably, they are not fully self-conscious yet.

⁴ To deal with the first construal of the obstacle, it could be argued that animals possess a primitive form of self-consciousness resting on non-conceptual first-person contents; see Bermúdez (2018).

of the first-person pronoun is not. Instead, the latter is a typical, rather than a necessary, condition. Though this is weaker, still a *prima facie* obstacle is raised; for if animals fail to fulfil the linguistic condition, they are unlikely to be self-conscious.

To overcome the obstacle, researchers have adopted the strategy of looking for non-linguistic behaviours that are functional analogues of the use of the first-person pronoun by adult humans, which in turn serve as measures of self-consciousness in animals. Three such measures have been operationalized, resulting in the following tests: mirror self-recognition, uncertainty monitoring and episodic memory. According to the first, animals are self-conscious just in case they recognize themselves in a mirror, as shown by the spontaneous exploration of their own bodies and body parts (see the review in Gallup & Anderson, 2020). According to the second, animals are self-conscious just in case they are aware of their own uncertainty as to the solution of meta-memory and perceptual discrimination tasks, as indicated by their largely opting out of difficult, rather than easy trials, when given the opportunity to do so (for details, see Smith et al., 2009; Beran, 2019). According to the third, animals are self-conscious just in case their behaviour shows them to remember past events in their own life; for instance, by the selective recovering of previously cached foods, depending on their decaying spans (for a review, see Clayton, 2017).⁵

According to the reported results, only some animal species pass the tests. On the face of it, this is a helpful outcome, for it demonstrates that a robust experimental methodology is at hand to address the question of animal self-consciousness; moreover, one that by-passes the linguistic obstacle being discussed. For what the tests (allegedly) do is provide evidence that, despite lacking the linguistic ability to use the first-person pronoun, some animal species possess a self-concept. Consequently, self-consciousness can be attributed to them.

So far, so good; but two further considerations threaten to cloud matters. First, as noted earlier, having a self-concept is a matter of distinguishing self and non-self, including the abilities to separate oneself from one's surroundings, both physical and social, and to discriminate between one's own and other people's mental states. Now, in so far as this is basic for individual survival, self-consciousness ought to be widely spread across the animal kingdom. However, the results from current tests tell a different story, as only some animal species manage to pass them. A way out is to argue that the tests do not measure the presence of a basic skill for survival. But then, what do they measure? And what is the point of researching into animal self-consciousness anyway, if the latter is so basic for survival that it must be acknowledged from the outset that evolutionarily successful species cannot but have it? Let us call this the puzzle of widespread self-consciousness. More formally, the puzzle rests on the tension between the following two theses:

T1. It is an empirically open question whether animals are self-conscious;

⁵ In contemporary animal research, the label "episodic-like memory" is often preferred, to avoid assuming without sufficient proof that conscious recollection is going on. The qualification is pertinent to the discussion here, but for the sake of concision, the simple label will be used throughout.

T2. Self-consciousness must be widespread, for having a self-concept is basic for survival.

Second, as already noted, current tests are valid to the extent that they are based on non-linguistic behavioural measures functionally modelled on the use of the first-person pronoun by typically self-conscious humans. The rationale behind this is that, when passing the tests, humans accompany their non-linguistic behaviour (say, pressing a button to opt out of difficult trials in uncertainty monitoring tests) with first-person utterances (e.g. “I pressed the button, because I was uncertain about my ability to answer correctly”). Therefore, similar results in button-pressing behaviour in animals must be a mark of self-consciousness, even in the absence of accompanying “I”-utterances; or so it is thought. Now, if in the human case button-pressing behaviour is proof of self-consciousness because of the ability to use the first-person pronoun, the latter cannot be a mere accompaniment, but rather must make a difference to the behavioural measure of self-consciousness. In turn, the following puzzle arises. Functionally analogous measures are required in the animal case, but if the requirement can be met in the absence of the linguistic ability to use the first-person pronoun, the latter must be mere accompaniment in the human case. Therefore, there is no reason to set the requirement in the first place. Yet, without it, animal performance by itself (e.g. the pressing of a button) could be explained in associative or other non-self-conscious terms. Let us call this the puzzle of non-linguistic self-consciousness. More formally, the puzzle arises because of the apparent conflict between the following theses:

T3. Settling the empirical question of animal self-consciousness requires a valid behavioural measure, in terms of functional analogy with typical self-conscious humans;

T4. There is no functionally-analogous measure of animal self-consciousness, in the absence of full-fledged linguistic abilities involving the first-person pronoun.

Now, the puzzles generated by the strategy to deal with the linguistic obstacle for the attribution of self-consciousness to animals must be solved in order to dispose of the obstacle. The proposal to be fleshed out next is that a distinction between two varieties of self-consciousness is the right tool for the job.

3 The varieties of self-consciousness

The question as to whether animals are self-conscious concerns the extension of a concept; so, to answer it, we need to have the concept plainly in view. To this effect, this section argues that there are two varieties of self-consciousness, as shown by the converging views of two contemporaneous twentieth-century philosophers, Sartre and Wittgenstein. For according to the reading of some of their key texts recommended here, both reject a purely epistemic conception of self-consciousness, proposing instead the existence of epistemic and non-epistemic varieties of

self-consciousness. Each author is discussed in turn, before bringing them together at the end of the section.⁶

3.1 Sartre on epistemic vs. ontological self-consciousness

To begin with, consider Sartre's treatment of the topic in *The transcendence of the Ego* (1936) and the introduction to *Being and nothingness* (1943), where he maintains that an epistemic conception of self-consciousness is contrary to the nature of consciousness, as revealed by phenomenological description. His argument can be reconstructed as the following *reductio*:

Sartre's argument from the nature of consciousness

1. Knowledge is a matter of entertaining (and thus having access to) a propositional content.
2. Consciousness is consciousness-of.
3. Acts of consciousness are self-conscious.
4. All self-consciousness is a matter of self-knowledge.
5. Therefore, for each and all of the acts of consciousness one undergoes, one consciously entertains the (propositional) content that one is undergoing them (i.e. a first-person propositional content).
6. For some acts of consciousness, when one undergoes them, one does not consciously entertain the propositional content that one is undergoing them – namely, first-order acts of consciousness, which are directed on to the world.
7. Yet, one's first-order acts of consciousness are self-conscious.
8. Therefore, not all self-consciousness is a matter of self-knowledge.

Premises (1) to (3) make explicit the background. On the one hand, (1) is a common assumption in epistemology, both in the classical definition of knowledge as justified true belief and in more recent post-Gettier refinements. On the other hand, (2) and (3) jointly characterize the notion of (acts of) consciousness pertinent to the argument – namely, intentional consciousness. Echoing Brentano, (2) could be rephrased as the claim that in conscious acts, there is something one is conscious of; i.e. their intentional object. Therefore, (3) is the idea that one's acts of intentional consciousness are self-consciously available to oneself.

Against this background, the aim of the argument is to correct a mistaken conception of the nature of self-consciousness pertaining to intentional consciousness. This is the epistemic conception made explicit in (4), i.e. the ability to know, of each and all of one's acts of consciousness, that one is undergoing them. Directly derived from it, (5) makes explicit the commitments of such a conception of self-consciousness, given the background premisses. In contrast, (6) presents a phenomenological finding, i.e. that in first-order acts of consciousness one's attention is totally taken up by some worldly intentional object or other, rather than by the relation between

⁶ For alternative readings of this material, see Longuenesse (2008) and Narboux (2018).

oneself and such an object. Premiss (7) is a corollary of (3), for first-order acts of consciousness are self-conscious. Therefore, to avoid contradiction between (5) on the one hand, and the conjunction of (6) and (7) on the other, the epistemic conception of self-consciousness is rejected in (8).

The argument is valid; but is it sound? Although the contradiction in the argument is traced to (4), the epistemic conception of self-consciousness is very intuitive indeed, at least at first sight. Therefore, before discarding it, several fixes could be attempted. Here, four such attempts are considered (and rebutted). Their common basis is that the argument goes through only if a particularly strong version of the epistemic conception of self-consciousness is read into premiss (4). According to it, self-consciousness is often thought to require (i) actual, (ii) conceptual or (iii) theoretical access to a propositional content, even (iv) a propositional content itself, although none of these are necessary conditions for self-consciousness. Therefore, to save the epistemic conception of self-consciousness, it is sufficient to weaken (4), by rejecting one or other of the assumed conditions. The point to be made now is that the fixes fail. Here are the details.

- i. *Actual vs. dispositional access.* It may be suggested that, contrary to the argument, self-consciousness does not require actual access to a first-person propositional content, but rather dispositional access to it. As a result, there is no contradiction in having one's attention actually taken up by the world in first-order acts of consciousness, whilst retaining dispositional access to the latter throughout. In this way, the conclusion of the argument is averted. The problem with this suggestion is that, if some form of the epistemic conception of self-consciousness is assumed from the outset, dispositional access must be a matter of a higher-order, self-reflecting consciousness, which in turn leads either to an infinite regress, or to a non-self-conscious (unconscious) starting point. According to Sartre, neither option saves the epistemic conception of self-consciousness, as they both fail to ground self-consciousness epistemically: either a never-ending move towards successive higher-order acts of consciousness is postulated; or a collapse into absurdity ensues, in so far as the ability to know, of (each of) one's acts of consciousness, that one is undergoing them turns out to be grounded in not knowing them (the unconscious).
- ii. *Conceptual vs. non-conceptual access (or content).* A different fix is that, contrary to what is suggested in the argument, self-consciousness need not require conceptual access to a first-person propositional content; rather (actual but) non-conceptual access is sufficient. (Alternatively put, self-consciousness is a matter of access to a non-conceptual first-person content.) Thus, as argued for by Proust (2009) on the basis of Gibson's (1979) ecological approach to visual perception, vision carries non-conceptual self-specifying information; i.e. information that is available to oneself in the absence of the conceptual abilities associated with language, rationality and self-reflection. As visual experience is an example of a first-order act of consciousness, the general conclusion of the argument is (apparently) avoided, by insisting that all that the epistemic conception of self-consciousness requires is the existence of self-specifying information, not a conceptual grasp of it. But Sartre's argument does not deny

that world-directed consciousness (vision included) carries self-specifying information, whether conceptual or not. Instead, its point is that one need not attend to (access) it; indeed, that phenomenological description shows that one does not actually attend to it, whilst engaged in first-order intentionality. Therefore, the emphasis on non-conceptual access (or content) does not avoid the conclusion of the argument.

- iii. *Theoretical vs. practical access.* Another proposal is that, contrary to a prominent reading of the argument, self-consciousness does not require theoretical access to a first-person content. Borrowing from Noë (2004), self-consciousness could be thought of as a form of sensorimotor knowledge, meaning not the ability to engage successfully with one's environment on the basis of one's bodily presence in the world, but rather one's grasp (understanding) of the sensorimotor contingencies that govern such an engagement, as manifested in one's expectations about what would happen if one's relation to the world were to change in specific ways. This practical knowledge differs from the theoretical, spectator-like entertaining of the corresponding counterfactual propositions. So, the suggestion is that the conclusion of Sartre's argument is escaped by taking the relevant notion of access practically. But Sartre's main point could be re-hashed as follows: it is one thing to expect certain things, it is quite another to reflect on them, as shown by phenomenology. So, without further ado, the theoretical vs. practical contrast does not save the epistemic conception of self-consciousness. If it is added that practical knowledge is to be thought of as tacit, a version of the difficulties mentioned in (i) applies once more.
- iv. *Knowing that vs. knowing how.* Finally, it could be maintained that, contrary to an explicit commitment of the argument, self-consciousness is not a matter of access to a first-person propositional content (knowing that), but rather a matter of exercising a first-person ability (knowing how). Accordingly, the problem with Sartre's argument is that it is guided by the wrong notion of knowledge (stated in premiss 1). But even if knowing how is accepted as a proper epistemic notion, the current suggestion will not save the epistemic conception of self-consciousness. For Sartre's distinction could be re-fashioned as the claim that it is one thing for first-order acts of consciousness to involve the ability for self-consciousness, and quite another for them to involve higher-order, self-reflecting abilities.

Therefore, the epistemic conception of self-consciousness cannot be salvaged by weakening premiss (4). But perhaps the fault lies with premiss (3)? Although for Sartre (3) is beyond dispute, it is quite common nowadays to accept that one can be conscious of something, whilst lacking awareness that one is so conscious; therefore, that there is consciousness without self-consciousness. Habitual agency is often cited as an example (though a similar analysis applies to non-habitual action, too). Thus, when one drives on automatic pilot (as it were), one adjusts one's performance to the traffic, despite being unaware of performing all the activities involved in doing so, from overtaking other vehicles to changing gears, braking and so on. Under such circumstances, one is conscious of the world, as

indicated by one's performance, despite not being aware of oneself and one's experiences as such. As a result, the notion of consciousness from which Sartre's argument takes off may seem implausibly strong. Moreover, adopting it as a starting point would appear to make this paper redundant; for having assumed from the outset that animals have mental states with intentional contents, the question whether they are self-conscious must be answered affirmatively, or so Sartre's starting point appears to entail.

In the example, drivers that are not reflectively aware of their driving are nonetheless aware of the world. Crucially now, in doing so they distinguish self from non-self, thereby implementing the ability for self-consciousness. Therefore, what the often-cited examples of habitual agency show is not that there is consciousness without self-consciousness, but rather that there is consciousness without reflective self-consciousness. And as the latter is just Sartre's point that in first-order acts of consciousness one's attention is totally taken up by some worldly intentional object or other, rather than oneself vis-à-vis such an object, far from undermining Sartre's starting point, the examples underscore its plausibility.

To sum up, neither premiss (3) nor premiss (4) is objectionable, which means that the soundness of Sartre's argument from the nature of consciousness has not been undermined. But what does its conclusion signify for the nature of self-consciousness? Together with the further thesis that one's first-order acts of consciousness can be known by oneself (i.e. when the propositional content that one is undergoing them is reflectively entertained), the argument must be taken to have shown that the self-consciousness of first-order consciousness (Sartre's unreflective consciousness) is a different phenomenon from self-knowledge (reflective consciousness). Importantly, in view of Sartre's avowed rejection of the primacy of knowledge in our understanding of intentional consciousness, unreflective and reflective varieties of self-consciousness must not be treated as two species of one genus, but rather as two different genera. In the terms used earlier, there are two varieties of self-consciousness, one epistemic, but not the other.

But what is self-consciousness, non-epistemically conceived? According to Sartre, the collapse of the epistemic conception of self-consciousness for first-order acts of consciousness shows that there must be an immediate (non-positional, non-thetic, pre-reflective) consciousness of consciousness, given that conscious acts are self-conscious. He glosses it thus: "it is one with the consciousness of which it is consciousness", which he further clarifies by saying that it is "a mode of existence"; moreover, "the only mode of existence that is possible for a consciousness of something" (1943, p. liv). So, in effect, his proposal amounts to replacing an epistemic with an ontological conception of self-consciousness, which could be summed up as follows: the immediate self-consciousness of first-order acts of consciousness is the *mode of being* of intentional consciousness. In order to illuminate what such an ontological conception of self-consciousness amounts to, let us turn to Wittgenstein.

3.2 Wittgenstein on epistemic vs. expressive self-consciousness

In his 1930s *Blue book* (1958, p. 66), Wittgenstein draws a distinction between two uses of the first-person pronoun. The passage occurs in a context where solipsism is being discussed, and is best read as criticizing an epistemic conception of self-consciousness. Here is a reconstruction of the argument:

Wittgenstein's argument from the use of "I"

9. Self-consciousness is typically manifested by the use of the first-person pronoun.
10. Self-consciousness is self-knowledge, i.e. a matter of access to, and recognition of, one's acts of consciousness (as one's own).
11. All self-conscious uses of the first-person pronoun rest on access to, and recognition of, one's acts of consciousness (as one's own).
12. Some uses of the first-person pronoun do not rest on access to, or recognition of, one's acts of consciousness (as one's own) – namely, the uses of the "I" as subject.
13. Yet, the uses of the "I" as subject are self-conscious.
14. Therefore, not all self-consciousness is a matter of self-knowledge.

Premiss (9) is a background assumption. Premiss (10) is the conception of self-consciousness targeted by the argument, and premiss (11) makes explicit a commitment following from it, given the background assumption. (12) is the key premiss in the argument; for in conjunction with (13), it amounts to contradicting (11). Therefore, to restore consistency, the epistemic conception of self-consciousness is rejected in (14). The alternative is to reject (9), but this would be very counter-intuitive indeed.

Given the pivotal role of (12), we must enquire whether it is warranted. Its warrant comes from what may be called the bipolarity of access (i.e. the idea that access to a state of affairs can be successful or not, or also correct or mistaken), together with the appreciation that uses of the "I" as subject do not exhibit this bipolarity. This is not the point that such uses are always correct (or never mistaken), in that the state of affairs being talked about is always successfully accessed. Rather, it is the point that the very idea of access, whether successful or not, does not apply.

To see this, consider that unsuccessful access would be a matter of misrecognition or misidentification, i.e. taking some acts of consciousness to be one's own when they are not, or vice versa. But for misrecognition to occur, provision has to be made for it. And this condition is not met in the use of the "I" as subject. This is not to claim that the use of the first-person pronoun rules out recognition and access across the board; but rather, that *some* uses of the first-person pronoun do. Thus, uses of the "I" as object are a matter of recognition and access (either successful or not); whereas uses of the "I" as subject are not.

Given that self-consciousness is typically manifested in the use of the first-person pronoun, what makes uses of the "I" as subject self-conscious, if not recognition and access? The answer is in two steps: first, uses of the "I" as subject

are intrinsically expressive acts; second, some such acts are intrinsically self-conscious. To expand on this, consider that uses of the “I” as subject are expressive acts (utterances). Thus, there is a use of (say) “I am in pain” in which it is a complaint; not a report about oneself based on access to a particular experience, recognized as one’s own. To be sure, there is also a use in which “I am in pain” is such a report; but this is just the duality of uses under consideration. A crucial point here is that using “I am in pain” to complain is not making an assertion about oneself; hence, talk of an expressive use of such utterances as “I am in pain” must be understood non-relationally. In other words, what makes the utterance a pain complaint is its content, which must not be thought of as a separate item (ontologically speaking) from the utterance itself, but rather as a formal, intrinsic quality of the utterance. In addition (the second step), some expressive acts (including utterances) intrinsically possess the quality of being done self-consciously (attentively or with full awareness), rather than half-consciously or distractedly. Thus, there is a use of “I am in pain” in which the latter is a self-conscious pain complaint, given the intrinsic quality of the utterance (i.e. its mode of being as the utterance it is).⁷

“I am in pain” is but one example among many; for the claims of the last paragraph can be extended to “I believe that p”, “I desire that q”, and other first-person, present-tense psychological utterances. This includes such proprioceptive and kinaesthetic claims as “I am sitting up” and “I am running fast now”, which are either claims about oneself based on the recognition of certain sensations, or the expression of certain bodily sensations and experiences. To appreciate the latter, it helps to notice that a psychological verb is implicit in the utterances, e.g. “(I feel as if/that) I am sitting up”. Summing up, a whole range of “I”-utterances exhibit the duality of (non-relationally) expressive and reporting uses under discussion. And more to the point here, some (non-relationally) expressive “I”-utterances possess the intrinsic quality of self-consciousness.

This means that it is misleading to gloss the distinction between the use of “I” as subject and the use of “I” as object as the contrast between psychological and bodily self-knowledge, respectively; that is, a contrast between two types of epistemic self-consciousness, as done by Evans (1982, Chap. 7). For one thing, uses of “I” as subject include what in the misleading gloss are taken to be (knowledge) claims about the state of one’s body, as in “I am sitting up”. For another, the distinction concerns different uses of what in the gloss are psychological utterances, like “I am in pain”, “I believe that p” and so on. Finally, it is not a distinction between two kinds of self-knowledge, but rather between two varieties of self-consciousness, only one of which is epistemic. What misleads people into thinking that it is a distinction between two kinds of self-knowledge is the view that all such “I”-utterances are self-ascriptions, for (as the thought goes) a claim is made (by oneself) about either one’s psychological or bodily properties and states. But such a view overlooks the (non-relationally) expressive use of the “I”-utterances under discussion.

⁷ For more on this reading of Wittgenstein, see García Rodríguez (2020).

The same oversight is the source of a common objection levelled against Wittgenstein's distinction – namely, that it is a *non sequitur* to conclude that the use of “I” as subject is not referential from the absence of recognition or identification, as done by Strawson (1959, Chap. 3), Shoemaker (1968) and Evans (1982, p. 218). This would be a fair objection to make if the “I”-utterances involved were self-ascriptions, but this is precisely what is being targeted by calling attention to the expressive use of “I”-utterances. Therefore, the *non sequitur* objection fails to hit the mark it is aimed at.

A further worry here is that expressive uses of “I am in pain” are equivalent to the report “Pain is going on”, i.e. the (first-order) reporting of some mental content, rather than the self-conscious report that one is undergoing pain. *Mutatis mutandis* for other expressive uses of first-person psychological sentences. Thus, when used expressively, “I believe that p” is equivalent to “Believing that p is going”; hence, the (first-order) reporting of some mental content, instead of the self-conscious report that one believes that p. Now, though talk of reporting may suggest so, the point being made cannot be that expressive uses involve the intention to convey some information to others, for one may be simply screaming with pain or acting in (linguistic) accordance with one's beliefs. Therefore, the point must be that others can find one's screams and behaviour informative; but rather than a problem, this is perfectly compatible with the expressive use of “I”-utterances defended here. Thus, screaming, whether accompanied or not by linguistic utterances, and the linguistic utterances alone, if used in the right circumstances, will be found by others to be informative about the mental condition of the utterer, precisely because the utterances play an expressive (i.e. plaintive) role. (*Mutatis mutandis* for “Believing that p is going on”.) Therefore, the worry under consideration fails to dent the preceding expressive analysis of “I”-utterances.

Summing up, two general conclusions follow. First, in addition to self-knowledge, there is an expressive variety of self-consciousness; to clarify, not two species of one genus, but rather two different genera. Accordingly, self-consciousness can involve reflective access and recognition, or can instead be a particular quality (or mode) of one's intentional engagement with the world. The former is based on a relation with oneself; whereas the latter captures a non-relational quality of (some) of the bodily-cum-behavioural features conforming such intentional engagement with the world. Both involve awareness of oneself and one's states as such, though in different ways; i.e. through reflection and action, respectively.

In support of the claim that, being expressive, some intentional engagement with the world is properly self-conscious, it is helpful to notice that linguistic complaints, no less than non-linguistic actions, can be self-conscious, in the stated sense. Thus, by quickly removing one's hand from a naked flame one shows awareness of one's painful experience as such, i.e. as both painful and one's own. Similarly, by screaming “That hurts!”, by complaining through an “I”-utterance, or indeed by doing either whilst removing one's hand from the flame, one shows awareness of one's painful experience as such.

Second, the notion of expression provides a gloss on Sartre's ontological conception of self-consciousness, for the self-conscious mode of existence of first-order acts of consciousness is an intrinsic quality of such acts, whether linguistic or not,

conceived in non-relationally expressive terms. The crux here is that the notion of expression targets the view that acts of consciousness are private events in an eerie medium, proposing instead that they are an intrinsic aspect of the (not only human) bodily-cum-behavioural mode of being in the world, linguistic utterances included. This fits the Sartrean idea that consciousness is a matter of transcendence rather than immanence, to use his terms; that is to say, a matter of how one is bodily and behaviourally in the world. So, the expressive gloss on Sartre's ontological conception of self-consciousness can now be summarized in the slogan: one is (intrinsically) self-consciously in the world (at least sometimes).

4 The varieties of animal self-consciousness

The previous section has shown that there are two varieties of self-consciousness: epistemic self-consciousness is a matter of being reflectively aware of oneself and one's mental states as one's own; whereas expressive self-consciousness is a mode of one's intentional, bodily-cum-behavioural engagement with the world. What this means for our main topic here is that the question as to whether some animal species are self-conscious branches into two, one for each variety of self-consciousness. Furthermore, different answers could be provided to each question, for some animals may fail to be epistemically self-conscious, whilst being expressively self-conscious. In fact, as will be claimed below, this is what the results from current experimental research show. But to do so, the significance of such research needs to be re-evaluated in the light of the distinction between epistemic and expressive varieties of self-consciousness. This is the assignment for the rest of the section.

As stated above, three experimental paradigms have been used to test what animal species are self-conscious: mirror self-recognition, uncertainty monitoring and episodic memory. They are thought to be valid tests of self-consciousness because they reveal which animals possess self-addressed abilities. Thus, in order to pass the mirror test, animals must treat the reflection seen in a mirror as their own (the self-directed response), rather than as another individual (the social response). Similarly, in order to pass the uncertainty monitoring test, animals must be aware of their own certainty or uncertainty regarding the successful exercise of the memory- and perception-based discrimination abilities needed to solve the primary task of the design. Finally, in order to pass the episodic memory test, animals must behave in the present as if remembering events (actions) in their own past.

Sometimes, the mentioned self-addressed abilities are construed in metacognitive terms; so, the claim is that if animals are to pass the tests, they must possess second-order abilities about their first-order, cognitive abilities. For instance, on this view, passing the uncertainty monitoring test involves the ability to track one's abilities for memory and perceptual discrimination; hence, a second-order ability. However, this is not a compulsory view to take. Thus, it has been argued that passing the test involves a first-order, relational ability – namely, to distinguish easy-for-one from difficult-for-one trials, depending on the different kinds of mental action afforded by the stimuli (Proust, 2009). Crucially, such a relational ability is self-addressed; therefore, self-consciousness is still at stake.

Regardless of how this dispute is eventually adjudicated, what bears emphasizing now is that the self-addressed abilities that (allegedly) hold everything together are being construed epistemically. In other words, what the designs do is provide evidence for the animals' ability to have access to, and recognize, themselves, their own cognitive states (say, uncertainty) or their own past, as such. Therefore, the designs are meant as tests for the presence of epistemic self-consciousness.

As also noted, the behavioural measures tested for are taken to be functionally analogous to the use of the first-person pronoun by typically self-conscious humans. Thus, spontaneous self-exploration in front of a mirror is the non-linguistic analogue of humans saying "That's me", upon recognizing their own image in a mirror, and acting accordingly. Similarly, opting out of a difficult trial in meta-memory and perceptual discrimination tasks is the non-linguistic analogue of the human acknowledgement of uncertainty contained in the utterance "I do not know what the right choice is" and subsequently pressing the opt-out button. Finally, the selective digging of previously cached foods, depending on their decaying spans, functionally resembles the linguistic utterance "I remember having some food that needs to be consumed promptly" whilst rummaging in the fridge. Now, what underwrites the claim of functional analogy is that all such "I"-utterances must be conceived in epistemic terms, i.e. as the result of a similar self-addressed ability.

To be sure, there is functional analogy between the animal and human cases; but the moot point is whether it is to be construed epistemically. For as noted in Section 3, not all "I"-utterances serve to report the results of a self-addressed ability; instead, some serve to express (intrinsically) one's own mind. Furthermore, in view of the fact that such a duality of reporting and expressive uses is linked to the epistemic and expressive varieties of self-consciousness respectively, adhering to the requirement of functional analogy means that current experimental designs could be testing for either epistemic or expressive self-consciousness in animals.

In this respect, a distinction must be drawn between the mirror test and the other two. It is clear that the social response to one's mirror image is a matter of failing to recognize oneself as such in the mirror; and conversely, that the self-directed response manifests recognition of oneself, or of one's body parts as one's own. Therefore, animals that pass the mirror test exhibit self-recognition, i.e. an epistemic ability. The crux here is that mirrors provide one with information about oneself that must be employed in order to pass the test. In the terms used earlier, the self-addressed abilities of animals that successfully pass the mirror test are abilities to use information about themselves. However, things are not so with the other tests; for though some self-information is available then to the experimental subjects, it plays no role in their passing the tests. Therefore, passing them is not a matter of exercising an epistemic ability.

To see this, consider that the utterances "I do not know what the right choice is" and "I remember having some food that needs to be consumed promptly" can be either reports about oneself or expressions of one's state of uncertainty and one's memories, respectively. If the former, a claim is made on the basis of some available information; for instance, noticing one's own hesitant behaviour or (more likely) a certain feeling, one may say "I do not know what the right choice is", as one may announce "They do not know, either" upon noticing other people's hesitant

demeanour. In saying “I do not know” under such circumstances, the focus is on oneself, and the self-information at one’s disposal is being used as the basis for the self-addressed claim. But one may utter the same sentence when the focus is elsewhere, i.e. on the matter at hand; in our example, on making the right choice. As the focus is not on oneself, the available self-information is not being used, as opposed to the information about the topic of one’s uncertainty. Under such different circumstances, one’s “I”-utterance is not a report (about oneself), but rather an expression of uncertainty (about the matter at hand), in a similar way in which shrugging one’s shoulders or pulling a certain a face (or both combined) are expressions of uncertainty. A matching distinction between making a report about oneself and expressing one’s mind applies to utterances of the form “I remember such-and-such”, when the latter concern events in one’s past. Therefore, the “I”-utterances that provide the model for current experimental research into animal self-consciousness are not always to be taken epistemically.

This is precisely the case in the uncertainty monitoring and episodic memory tests. Then, the animals’ focus is not on themselves, but on the primary task at hand. This follows from the fact that, in the uncertainty monitoring tests, they have been trained to respond selectively to certain stimuli in order to obtain a reward, and in the key trials they are intent in doing well to secure the reward. Similarly, in the episodic memory tests, the animals are focused on recovering previously cached food. So, despite the fact that some self-information is available, to themselves as much as to others, the animals are not using it; for their focus is elsewhere. This preserves the functional analogy between the animal and human cases, but outside an epistemic framework. Therefore, the conclusion must be that what the designs do is provide evidence of expressive self-consciousness.

To elaborate: as argued for above, expressive self-consciousness is (intrinsically) a matter of one’s mode of intentional, bodily-cum-behavioural engagement with the world, in so far as the latter is neither half-conscious nor distracted. Therefore, the claim being made now is that a similar intrinsic quality belongs to the animals’ engagement with the world when performing certain actions (i.e. pressing the opt-out button or digging selectively) in response to the primary tasks set in the uncertainty monitoring and episodic memory tests. This is why the animals that pass the tests exhibit expressive self-consciousness.

This is not to say that any correct behavioural episode amounts to passing the tests (or exhibiting expressive self-consciousness). For one thing, the episode might a happy accident, if not repeated on a sufficient number of occasions or in related though varied conditions. For another, each lucky episode might have a distracted or half-conscious quality about it. Under such synchronic and diachronic circumstances, the animal’s mode of engagement with the world will not qualify as self-conscious, much like the behavioural episodes of young children and some adult humans, “I”-utterances included, fail to qualify as self-conscious under similar circumstances. Therefore, in order to attribute expressive self-consciousness to animals, care must be taken not to rush to conclusions unsupported by sufficient evidence, including the spontaneous transfer of correct behaviour to new conditions, without the need for intensive re-training. As only such temporally extended patterns count as passing the tests, only they provide evidence of expressive self-consciousness.

Nonetheless, animals failing the tests in the specified ways nevertheless exhibit consciousness of the world under some aspect, though not the aspects pertinent to the tests. Therefore, failing the tests is not evidence of absence of self-consciousness; at most, it is evidence of absence of epistemic self-consciousness. If this is thought to beg the question whether animals are self-conscious, it bears emphasizing (in line with a point made in Section 3 above) that as long as animals are agreed to engage intentionally with the world, they must be able to distinguish between self and non-self, which is what self-consciousness amounts to. And conversely, that unless animals are counter-intuitively denied intentional mental states (particularly the primates that have been the focus of attention here), they must be self-conscious.

Against this backdrop, what the expressive case for animal self-consciousness does is remove the negative influence that the epistemic variety of self-consciousness has on our understanding of animal self-consciousness; and this for two reasons. First, it encourages the conclusion that animals are not self-conscious from their failing the tests of epistemic self-consciousness, thereby overlooking the expressive variety. Second, it fosters the suspicion (or the outright charge) that the appeal to the expressive variety, thought of as an intrinsic quality of bodily-cum-behavioural intentional consciousness, begs the question of animal self-consciousness, thereby neglecting that the existence of a first-person perspective goes hand-in-hand with intentional engagement with the world, also in animals.

Overall, it is an unduly restrictive conception of self-consciousness, construed in terms of self-access and self-recognition, what hinders our clarity concerning the matter of animal self-consciousness. It is also what underlies the two puzzles mentioned earlier in the paper. Therefore, to round off the expressive case for animal self-consciousness, the next section is devoted to spelling out the solution to the puzzles.

5 Two puzzles solved

As stated, two puzzles threaten to undermine the strategy endorsed by current researchers to overcome the linguistic obstacle for the attribution of self-consciousness to animals. But they do so in different ways: the puzzle of widespread self-consciousness calls into question the validity of the experimental results; whereas the puzzle of non-linguistic self-consciousness compromises the validity of the designs as a whole. This section shows that the distinction between epistemic and expressive self-consciousness suffices to defuse the threats.

To begin with the former puzzle, if self-consciousness requires possession of a self-concept, and the latter involves the ability (basic in survival terms) to distinguish oneself from one's surroundings, it is surprising that only some animal species pass the tests for self-consciousness. To see this, consider the mirror test. As noted above, passing the test requires recognition of one's body or body parts as such.

Now, according to one leading assessment of the available evidence (Anderson & Gallup, 2015), only the great apes pass the test, for only they use mirrors in spontaneous self-exploration of their bodies. By contrast, other primate species fail the test, as shown by their social response to the mirror image. To be sure, there is room for disagreement here, as monkeys have been shown to pass an alternative mirror test, in which the coloured mark on the animals' forehead has been replaced by an implant (Rajala et al., 2010).⁸ But Anderson and Gallup reject this conclusion, as it is based on intensive training of the animals, rather than on spontaneous behaviour, and therefore is subject to a deflationary explanation in terms of associative learning. Now, if only spontaneous behaviour is accepted as a valid measure, a version of the puzzle of widespread self-consciousness arises. For, on the basis of the evidence available, it can be alleged that, as only the great apes pass the test, they alone possess a self-concept; whereas the ability to distinguish oneself from one's surroundings is so basic for survival that all primates should possess a self-concept, including the species that do not succeed at the test.

Consider now the following reported evidence (Gallup & Anderson, 2020): primates that fail the test, use mirrors to investigate their physical surroundings, particularly places they would not have access to otherwise (the environmental response). The latter shows that they are alert to their physical surroundings as distinct from themselves. Something similar is shown by the social response. In their mistake, the animals take their mirror image for another individual (a competitor perhaps), and respond accordingly (say, aggressively). But in doing so, they show to be alert to their social surroundings as distinct from themselves. Therefore, the primate species that fail the test nonetheless exhibit possession of a self-concept by their use of mirrors in the environmental and social responses.

How can all this evidence be put together? Here is one proposal: although only some primate species show self-recognition and therefore epistemic self-consciousness (according to one leading assessment of the evidence), all primates exhibit expressive self-consciousness. For, by their use of mirrors, primates exercise the ability to distinguish themselves from their surroundings, both physical and social, thereby exhibiting a self-concept. In support of the latter, consider that their use of mirrors is a functional analogue of similar behaviour by typically self-conscious humans when they engage with their surroundings; behaviour that may be accompanied by such utterances as "I can see it now". Notwithstanding the fact that the utterance could serve a different purpose in other circumstances, as the humans' focus is then on the world, the utterance should be thought of as an expression of their own perceptual experience (what can be seen), rather than a report about the fact that they are having such an experience. Moreover, there is no reason to rule out that the utterance possesses (or may possess) the intrinsic quality that characterizes self-conscious engagement with the world; hence, that under the circumstances the humans exhibit expressive self-consciousness, and therefore possession of a self-concept. Now, by parity of functional role, primates' use of mirrors to explore their

⁸ Other worries, e.g. that vision is not the leading sense modality of many non-primate species, will be glossed over; for to keep the discussion manageable, only primates are considered.

surroundings should also be thought of as an expression of their self-concept, for their behaviour intrinsically expresses their perceptual and social way of being in the world.

In turn, this makes available both a diagnosis and a remedy for the puzzle of widespread self-consciousness. According to the diagnosis, the puzzle arises if the epistemic conception of self-consciousness alone guides our understanding of the research into animal self-consciousness. The remedy is to acknowledge the expressive variety of self-consciousness and give it its proper place in current research. Furthermore, diagnosis and remedy apply to all three experimental designs, not only the mirror test. For if the uncertainty monitoring and episodic memory tests are (mistakenly) interpreted under the aegis of the epistemic conception of self-consciousness, a version of the puzzle arises; particularly, if there are doubts over the success of some animal species at the tests, as is the case with pigeons (Shettleworth & Sutton, 2006) and rats (Crystal, 2019). As shown in Section 4, the mistake lies in that the results from both tests are best understood as proof of expressive self-consciousness. This includes negative results, which rather than being evidence of absence of self-consciousness, indicate the boundaries of the species' (first-order) intentional engagement with the world. The bottom line is that current methodology is not threatened by the puzzle of widespread self-consciousness, as long as both varieties of self-consciousness are kept in view.

Turning now to the puzzle of non-linguistic self-consciousness, it is the validity of the experimental designs as a whole that is under threat. For if the only available measure of animal self-consciousness is non-linguistic behaviour that is functionally analogous to the behaviour of typically self-conscious humans that pass the tests, then as human behaviour is accepted as proof of self-consciousness because it can be accompanied by use of the first-person pronoun, a dilemma ensues. For either functional analogy can be achieved in the absence of the first-person pronoun, which makes the required analogy with the linguistic case redundant; or else the requirement is fine, but in the absence of the first-person pronoun functional analogy cannot be guaranteed, as an alternative non-self-conscious explanation cannot be ruled out. As will be shown forthwith, the key to safeguarding current methodology lies in a proper appraisal of what humans are doing when they use the first-person pronoun.

Consider the uncertainty monitoring test. When humans press the opt-out button in difficult trials, they add that they did so because they were uncertain about their ability to make the right choice. It is tempting to gloss this as follows: humans have access to, and recognize, their own uncertainty, and as a result press the opt-out button. As noted earlier, the gloss can be construed either metacognitively or as the direct pick-up of the relational, self-specifying information present in the stimulus. But either way, some information about oneself (i.e. one's uncertainty) is tracked. On this view, what sustains the attribution of self-consciousness to the humans that by pressing the opt-out button pass the test is their self-addressed epistemic abilities, as typically manifested by their "I"-utterances on such occasions. By the same token, the test provides evidence of epistemic self-consciousness. In turn, this gives rise to the puzzle under discussion, and in the absence of the linguistic ability to use the first-person pronoun, the animals' analogous button-pressing behaviour could be explained otherwise than by the awareness of their own uncertainty. For instance, it

could be explained in terms of different strengths of first-order beliefs and desires (Carruthers, 2008), or other non-self-addressed abilities (Crystal & Foote, 2009). Under such circumstances, though the animals behave adequately, in that they largely press the opt-out button in difficult, but not in easy trials, no valid measure of self-consciousness has been provided by the design.

Underlying this sceptical conclusion is the idea that, with their behaviour, animals exhibit either self-addressed epistemic abilities or non-self-conscious first-order abilities. But this overlooks the expressive variety of self-consciousness. Consider the human case, once more. When humans state that they press the opt-out button because they are uncertain, their “I”-utterances are best understood not as reports, but as expressions of their state of uncertainty. This is so for the familiar reason that their attention is on the primary task at hand, rather than on themselves performing the task. On such occasions, the “I”-utterances belong with the button-pressing behaviour, contributing to the (intrinsically) self-conscious quality of the latter. In other words, it is the combination of “I”-utterances and button-pressing behaviour that carries the self-conscious quality of an act performed in full awareness of one’s surroundings (here, the trials and stimuli being presented with).

What this means is that preserving functional analogy between the animal and human cases involves identifying a behavioural measure of expressive self-consciousness in the absence of the linguistic ability to use the first-person pronoun. At this point, it is helpful to remember that although self-consciousness is typically manifested by the use of the first-person pronoun, the latter is not a necessary requirement: other linguistic and non-linguistic devices are often sufficient. Thus, uncertainty can be expressed by non-linguistic hesitant behaviour; for instance, taking extra time before choosing, reversing one’s decision if possible, looking pensive, and so on. In fact, such combined tell-tale features are both an intrinsic quality of the behaviour at hand and reliably identified in ordinary circumstances. (Reliably, rather than infallibly, because fallibility is a background, yet manageable invariant). Therefore, what preserves functional analogy between the human and animal cases is the existence of a behavioural measure that is intrinsically expressive of uncertainty in the experimental subjects. The main difficulty here is spelling out the exact contour of that measure in the species under investigation, for it need not coincide exactly with the human combination of tell-tale features. But this need not be an unsurmountable obstacle, especially in the primate species that are our current focus, given their anatomic, physiological and evolutionary proximity to humans.

As above, the foregoing makes available a diagnosis of the puzzle of non-linguistic self-consciousness, and a remedy for it – namely, that the puzzle arises if current methodology is guided by the epistemic variety of self-consciousness alone, and disappears if the expressive variety is acknowledged. Furthermore, this conclusion applies not only to the uncertainty monitoring test, but to the episodic memory test as well. Thus, if the latter is guided by the epistemic conception of self-consciousness, in the absence of the linguistic ability to use the first-person pronoun, the doubt may arise as to whether recovering previously cached food is ever a valid measure of self-consciousness. But the doubt dissipates if an expressive variety of self-consciousness is adopted, instead; for a valid measure is made available – namely, the intrinsic quality of the recovering behaviour.

The former also applies to the environmental and social responses of the subjects in the mirror self-recognition tests. For, on those occasions, their behaviour possesses an intrinsic quality to the effect that, even in the absence of the linguistic ability to use the first-person pronoun, doubts about the expression of a self-concept do not arise. After all, when the animals' behavioural episodes are not a happy accident, but are repeated in a number of different and varied occasions, their behaviour exhibits acknowledgement of their physical and social surroundings as distinct from themselves.

Therefore, although it may seem that, to solve the puzzles, one in each of the pairs of conflicting theses underlying them must be abandoned, the proposal of this paper is that, as soon as the existence of two varieties of self-consciousness is acknowledged, the conflicts are revealed as apparent. On the one hand, the puzzle of widespread self-consciousness is solved by showing that

T1. It is an empirically open question whether animals are self-conscious

is about the epistemic variety (which need not be widespread nor basic for survival); whereas

T2. Self-consciousness must be widespread, for having a self-concept is basic for survival

is about the expressive variety (which is both widespread and basic for survival, on account of the fact that animals engage in intentional, bodily-cum-behavioural relations with their surroundings).

On the other hand, the puzzle of non-linguistic self-consciousness is solved by defusing the challenge allegedly posed to

T3. Settling the empirical question of animal self-consciousness requires a valid behavioural measure, in terms of functional analogy with typical self-conscious humans

by

T4. There is no functionally-analogous measure of animal self-consciousness, in the absence of full-fledged linguistic abilities involving the first-person pronoun.

For the challenge is based on an understanding of the "I"-utterances that provide the model for animal self-consciousness research that fails to acknowledge the expressive variety of self-consciousness.

So, to return to the question posed at the beginning of the paper: are some animal species self-conscious, perhaps the primates widely used in contemporary experimental research? What has been shown is, first, that our clarity in this area is clouded by an unduly restrictive interest in epistemic (i.e. reflective) self-consciousness; second, that there are epistemic and expressive varieties of self-consciousness;

and third, that there is evidence of both in animals. In this way, an expressive case has been made for the attribution of self-consciousness to animals.

Acknowledgements The constructive comments received from the reviewers appointed by the journal are gratefully acknowledged.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

The author has no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, J. R., & Gallup, G. (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates*, *56*, 317–326.
- Bekoff, M. (2003). Consciousness and self in animals: some reflections. *Zygon*, *38*(2), 229–245.
- Bekoff, M., & Sherman, P. W. (2004). Reflections on animal selves. *Trends in Ecology and Evolution*, *19*(4), 176–180.
- Beran, M. J. (2019). Animal metacognition: a decade of progress, problems, and the development of new prospects. *Animal Behavior and Cognition*, *6*(4), 223–229.
- Bermúdez, J. (2018). *The bodily self*. The MIT Press.
- Birch, J., Alexandra, K., Schnell, & Nicola, S. C. (2020). Dimensions of animal consciousness. *Trends in cognitive sciences*, *24*(10), 789–801.
- Carruthers, P. (2008). Meta-cognition in animals: a skeptical look. *Mind and Language*, *23*, 58–89.
- Cassam, Q. (1994). Introduction. In *Self-knowledge* (pp. 1–18). OUP.
- Clayton, N. S. (2017). Episodic-like memory and mental time travel in animals. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: perception, learning, and cognition* (pp. 227–243). American Psychological Association.
- Crystal, J. D. (2019). Comparative approaches to metacognition: prospects, problems, and the future. *Animal Behavior and Cognition*, *6*(4), 254–261.
- Crystal, J. D., & Foote, A. L. (2009). Meta-cognition in animals. *Comparative Cognition and Behavior Reviews*, *4*, 1–16.
- de Waal, F. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLoS Biology*, *17*(2), e3000112.
- DeGrazia, D. (2009). Self-awareness in animals. In R. W. Lurz (Ed.), *The philosophy of animal minds* (pp. 201–217). CUP.
- Evans, G. (1982). *The varieties of reference*. Clarendon Press.
- Gallup, G., & Anderson, J. R. (2020). Self-recognition in animals: where do we stand 50 years later? Lessons from cleaner wrasse and other species. *Psychology of Consciousness: Theory, Research, and Practice*, *7*(1), 46–58.
- García Rodríguez, Á. (2020). A Wittgensteinian view of mind and self-knowledge. *Philosophia*, *48*(3), 993–1013

- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Heyes, C. M. (1994). Reflections on self-recognition in primates. *Animal Behavior*, 47, 909–919.
- Longuenesse, B. (2008). Self-consciousness and self-reference: Sartre and Wittgenstein. *European Journal of Philosophy*, 16(1), 1–21.
- Narbox, J. P. (2018). Is self-consciousness consciousness of one's self?. In O. Kuusela, M. Ometiță, & T. Uçan (Eds.), *Wittgenstein and phenomenology*, (pp. 197–247). Routledge.
- Noë, A. (2004). *Action in perception*. The MIT Press.
- Perry, J. (1979). The problem of the essential indexical. *Nous*, 13, 3–21.
- Proust, J. (2009). The representational basis of brute metacognition: a proposal. In R. W. Lurz (Ed.), *The philosophy of animal minds* (pp. 165–183). CUP.
- Rajala, A. Z., Reininger, K. R., Lancaster, K. M., & Populin, L. C. (2010). Rhesus monkeys (*Macaca mulatta*) do recognize themselves in the mirror: implications for the evolution of self-recognition. *PLoS One*, 5(9), e12865.
- Sartre, J. P. (1936/2004). *The transcendence of the Ego*. Routledge: Translation by Andrew Brown.
- Sartre, J. P. (1943/1956). *Being and nothingness*. Washington Square Press. Translation by Hazel E. Barnes.
- Shettleworth, S. J., & Sutton, J. E. (2006). Do animals know what they know?. In S. Hurley, & M. Nudds (Eds.), *Rational animals* (pp. 235–246). OUP.
- Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, 65/19, 555–567.
- Smith, J. D, Beran, M. J., Couchman, J. J., Coutinho, M. V. C., & Boomer, J. B. (2009). Animal metacognition: problems and prospects. *Comparative Cognition and Behavior Reviews*, 4, 40–53.
- Strawson, P. F. (1959). *Individuals*. Methuen.
- Wittgenstein, L. (1958). *The blue and brown books*. Blackwell.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.