

On the Possibility of Self-Awareness*

Markus Werning

To discuss the issue of consciousness and awareness, philosophers have coined the notion of a zombie. By definition, zombies are like us in all but one respect: they don't have any states of consciousness. They aren't conscious of anything that goes on in the external world or inside their bodies. They are subjects neither to perceptions, nor to hallucinations. They feel neither pain, nor hunger, feel neither happy, nor sad. We aren't zombies and we are aware that we aren't. Each of us takes that for granted. Some philosophers, most prominently Descartes, even hold that nothing is as certain as the existence of the states of consciousness of one's own. The fact that we know that we aren't zombies, nevertheless, poses a philosophical problem. As Dretske (2003, p.1) remarked recently: "There is absolutely nothing you are aware of, external or internal, that tells you that, unlike a zombie, you are aware of it. Or, indeed, aware of anything at all." How do we know, then, that we aren't zombies? How do we know that we are conscious of anything at all?

The problem roots in what is often called the transparency of consciousness. This issue was first advanced by G. E. Moore (1903) and has since been reformulated in various ways.¹ For want of a more felicitous notion, I will continue to use the notion "transparency", although I do not want to imply that states of consciousness are diaphanous in any other but a metaphorical sense.² What I have in mind is this: If you are conscious of a lion hunting a zebra, being in that state is for you just as if you were in an environment with a lion hunting a zebra. It is, particularly, not like watching *Out of Africa* in the movies or reading Hemmingway's *The Snows of Kilimanjaro*. This is so despite the fact that there are plenty of pictorial and propositional theories of consciousness on the market. If you are conscious of a lion hunting a zebra you will feel, behave and form beliefs as if there were a lion hunting a zebra in that region of space. This holds whether or not there really is a lion, a zebra or a hunt, i.e., whether or not your perception is veridical or whether you are hallucinating.

The transparency of consciousness seems to imply that the subject when being conscious of the lion does not have any awareness of her being conscious of the lion. The argument has the form of a dilemma: Being conscious of a lion and being in an environment with a lion are for the subject either the same or not the same. If both are the same for her, she doesn't have any empirical reasons to believe that she is conscious of the lion. All that is for her can be explained on the basis that there is a lion (and a zebra and

* I am grateful to Fred Dretske who gave me the opportunity to discuss the issues involved in this paper in greater depth on several occasions. I am thankful for helpful comments to Brian McLaughlin, Robert Hopkins, Alex Burri, Christian Beyer, Wolfgang Huemer, and Gerhard Schurz. I owe many of the insights on autism and mindreading to Alan Leslie and Stephen Stich.

¹ See also Harman (1990).

² Perhaps "direct realism" might be as good a label as "transparency of consciousness".

a hunt and all the other external objects and events she is conscious of). From her epistemic perspective, there is no need to additionally postulate a state of consciousness, be it a representation, or be it a cluster of internal qualia. If being in an environment with a lion and having consciousness of the lion are not the same for her, she might have empirical evidence for an intermediate representational or phenomenal state – she might be aware of that intermediate state – but, then, being in that state fails to be for her as if she were in an environment with a lion. It rather is for her as if she were in a world with a representation of a lion or with a lion-like cluster of qualia.

The problem of a subject knowing that (s)he is not a zombie, thus, consists of two – as it seems – mutually exclusive propositions:

(1) *Self-awareness*

Sometimes, if a subject is conscious of X, she is aware that she is conscious of X.

(2) *Transparency of consciousness*

If a subject is conscious of X, being in that state always is for the subject just as if she were in an environment with X.

By stipulation I am using “conscious” not as a success word. I want the term “consciousness” to comprise, at least, perception and hallucination. For, both attitudes are indistinguishable from a subjective point of view and this is the point of view I am taking right now.

The transparency thesis apparently implies that we aren’t aware of any states of consciousness when we are conscious of something. Self-awareness of this sort, however, seems to be required in order to know that one is not a zombie. If consciousness is transparent – one may conclude –, we do not know that we aren’t zombies. This, however, is an absurd consequence.³ In this paper, I will suggest a way of reconciling the capacity of self-awareness with the transparency of consciousness.

³ If the transparency thesis leads to an absurd consequence, why not simply deny that perception and hallucination are transparent? This is in fact what many philosophers have done implicitly or explicitly. Instead of taking the external objects and events subjects are conscious of as prior, they sought to logically reconstruct them from internal entities. Carnap (1928), e.g., started his reconstruction from sensations. According to him, being conscious of a table is not just like being in a world with a table. It rather is like being in a world with sensations, classes of sensations, classes of classes of sensations, etc. From this point of view, the table is nothing but a hypothetical postulate. Others, e.g. Chisholm (1966), start off with evident thoughts. Being conscious of a table is supposed to be not just like being in an environment with a table, but rather like being in a world with the thought that there is a table. That the table is present, still, is a consequence of some inference. The transparency thesis, to be sure, does not deny that some intermediate states internal to the subject are causally responsible for the fact that the perception or hallucination of something is for the subject as if she were in a world with that thing. It just implies that the subject isn’t aware of these internal states and that it consequently cannot use reports about them as premises of inferences. Despite the phalanx of its opponents, the transparency thesis is so compelling because perception tells us much more about the structure of the world than it tells us about the structure of our minds, to say nothing about our brains. The best evidence, probably, is

Proprioception

Is there anything you are conscious of that tells you that you are conscious of it? Or, to put the question differently, have you ever noticed anything that was such that it can impossibly exist in a mind-independent way? If so, you would be licensed to infer that there are minds and that what you noticed was nothing but a content of your consciousness. Dretske argues convincingly that the existence of bodily sensations like pain does not indicate to you in a non-question-begging way that you have consciousness. Here's a somewhat different argument that leads to the same conclusion: There are two ways of thinking about pain. First, pain can be conceived of as a state that necessarily involves consciousness. Then, pain is itself a state of consciousness rather than a content of consciousness. Having a toothache, according to this view, is being conscious of a sore tooth. We again face the dilemma: Having a toothache and being in a world with a sore tooth are for the subject either the same or not the same. If they are the same for the subject, there is no reason to postulate anything over and above the tooth. If having a toothache and being in a world with a sore tooth, on the other hand, are not the same for the subject, the transparency thesis is violated. According to the second way of thinking about pain, it does not necessarily involve consciousness. It simply is a state of the body of which you occasionally become conscious when you *feel* pain. Then, however, noticing pain does not license any inference to the existence of a mind because pain, under this conception, can possibly exist independent from minds.

Hallucinations

Maybe you happen to hallucinate about weird things that are very unlikely to exist in reality. Even if you did, what is improbable since hallucinations are rare, it would not help you with the problem of how to know that you are not a zombie. Recall that we have not collected any evidence for the existence of the minds of our own, yet. Suppose somebody has a hallucination of a unicorn. What would be more rational for him: to accept that he is in a world with a unicorn or to postulate that the unicorn is nothing but the content of his consciousness? The first option would be more rational, I suppose. Although unicorns are weird objects, they still are objects. Insofar, they resemble the objects he else has been conscious of more than minds do. The postulation of unicorns, therefore, is ontologically much less extravagant than the postulation of minds would be. This consideration can be generalized for all hallucinated objects and events however weird they are, provided that even hallucinations never are logically inconsistent.

the history of the philosophy of mind. The proposals what the mind is build of (sentence-like entities, mental pictures, ideas, qualia?) are much more controversial than proposals of what the world is built of: objects and events with their properties and relations.

Introspection

Our perspective towards consciousness has still been too narrow to answer the question how we know that we aren't zombies. What else are we to look for? A widespread presupposition is the following: If a being has states of consciousness and if it is sufficiently rational, intelligent and has sufficiently strong conceptual resources, it will at some point discover that it has states of consciousness.⁴ That apes, monkeys and perhaps other non-human animals lack self-awareness, although they possess the capacity of perception and proprioception and are not immune against hallucinations, is explained by a deficit of rationality, intelligence or conceptual resources. The preceding considerations, however, suggest that nothing about perception, proprioception and hallucination provides a justification even for a perfectly rational and intelligent subject to postulate states of consciousness.

To explain our capability of self-awareness, we have to look for a capacity that satisfies three conditions: (i) The capacity is not implicit in that of perception, proprioception, or hallucination; (ii) it is not as general as rationality, intelligence, or conceptual resources; and (iii) it is sufficient to occasionally become self-aware when the subject comes to have a state of consciousness.

Often the faculty of introspection is postulated at this point. Supposedly, animals without self-awareness do not possess this faculty. It is not as general as rationality and should suffice for self-awareness. Unfortunately, the states of perception, of proprioception and of hallucination do not, by introspection, constitute any additional objects you are conscious of if the transparency thesis is true. If the perception of a lion were itself an object of introspective consciousness, being in the perceptive state and the introspective state at the same time would be for the subject as if she were in a world with a lion and with a perception of a lion. According to the transparency thesis, the latter – the perception of a lion – is for a subject just as if she were in a world with a lion. It follows that being in the combined – perceptive plus introspective – state is for the subject as if she were in a world with a lion and with the same or another lion. Since the subject is obviously not conscious of two lions, being in the states of perception and introspection at the same time is for the subject just as if she were in a world with one lion. It hence is just like being in the state of perception without being in the state of introspection. Introspection adds nothing to perception. For the same reason it adds nothing to hallucination and proprioception.

Object binding

Every instance of perception is an instance of object or event consciousness. This does not preclude that you sometimes perceive properties or facts. Whenever you perceive a property, you will, nevertheless, be conscious of this property as instantiated by an individual, i.e., by an object or event. Whenever you perceive a fact, you will be conscious

⁴ This presupposition is implicit, e.g., in Sellar's (1953) Jones myth.

that some individual has some property or stands in some relation to other individuals. The issue of what cognitive scientists call object (or event) binding is essential to perception.⁵ In perception you never are conscious of a property unless this property is bound to an object or event. The same is true for hallucination. Otherwise hallucinations could be distinguished from perceptions on the basis of their contents alone.

The solution to the zombie problem, I will argue, is that in imagination this is not the case. In imagination, you are conscious of properties or sequences of properties that are not bound to objects or events. This explains the indeterminate character of imagined contents. Since in all possible worlds properties are bound to objects or events, the content of imagination allows you to infer that you have a mind. Unbound properties can impossibly be anything other than contents of consciousness.

Recall that it is at the very core of consciousness and mental states in general to be intentional in the sense that, if an intentional state is directed towards the target X, X need not exist. One can now turn the tables: If you notice an X that can impossibly exist, the best explanation is that X is the content of an intentional state. If you notice something that can impossibly exist, you are licensed to infer that there are intentional states, hence minds. In imagination you notice unbound properties, which can impossibly exist, therefore you'd better postulate minds.

Before I turn to imagination in more detail, let me go through some arguments against the claim that object or event consciousness is essential to perception and hallucination. To justify that the perception of a property does not presuppose consciousness of an object, Dretske (1999) mentions a person who perceives the color of Tim's tie, which is blue, without being aware of the tie. The person sees another object, Jack's shirt, of exactly the same color and somebody tells him: "Jack's shirt has the color of Tim's tie." What Dretske ignores, however, is that the person still is conscious of this very shade of blue as instantiated by *some* object: not by Tim's tie, but by Jack's shirt. So even here, property perception presupposes object consciousness.

Dretske (1999) also mentions that people sometimes perceive movements without being conscious of any objects that move. To make this an argument for the claim that perception of properties is possible without consciousness of an object, you need the further premise that movements are properties of objects. It can, however, be argued that movements and the like are best treated as events, which, logically speaking, are taken to be individuals. Whenever an object is moving the object stands in the relation *being subject of* to the movement. It's not the relation of instantiation that holds between the object and the movement.⁶

Although property perception always presupposes consciousness of an object, it does not always presuppose the perception of an object. This is because "to perceive" is a success word whereas "to be conscious of", as I introduced it, is not. Property percep-

⁵ For the issue of object binding see Treisman (1996). For the less frequently studied issue of event binding see Hommel et al. (2001).

⁶ That events have to be treated as individuals rather than properties is argued explicitly by Pianesi and Varzi (2000).

tion without object perception often takes place in cases of illusion. You stare at a hologram and are conscious of a brown horse. Because the horse is illusionary, it would be false to say that you perceive a horse. The color brown, however, is something you do perceive.

Dretske (1999) is externalist with respect to object consciousness. He assumes that some causal relation between the object one is conscious of and the state of consciousness is to hold in case of object consciousness. Consequently, hallucinations cannot be cases of object consciousness. For Dretske, they merely are cases of property consciousness. This however would restrict the complexity of hallucinated contents. Consider three sample cases:

- (1) The hallucination of two pumpkins as opposed to the hallucination of one pumpkin.
- (2) The hallucination of a red square in a green circle as opposed to the hallucination of a red circle in a green square.
- (3) The hallucination of a slowly rotating quickly flying ball as opposed to the hallucination of a quickly rotating slowly flying ball.

The content of these hallucination is logically too complex to be analyzed as a property or a sequence of properties. At some point one has to introduce, within the scope of the representational operator, either a singular term (e.g., an indexical like “this” or “that”) or a variable bound by an existential quantifier. This, however, would imply that the sample cases are instances of object consciousness. Whether hallucinations that complex really occur is an empirical question, of course.

Imagination

We know from psychopathology that hallucinations typically are transparent. When a schizophrenic patient hallucinates about a voice, it really is for him as if he were in a world with a voice. Consequently, schizophrenics typically try to explain the occurrence of the voice by forming hypotheses about the world. Since they hear the voice without seeing a person speak to them, they, e.g., infer: God is talking to me; the CIA are tailing me and giving instruction to me through a loudspeaker; etc. What makes the hallucinated objects and events of schizophrenics so much more “real” than our auditory or visual imaginations of which we are conscious when we are thinking or day-dreaming? The contents of hallucinations, I suppose, must be determinate with respect to properties these object or events typically are perceived to have. A hallucinated voice must, e.g., have a certain sound, a certain intonation, a certain loudness and last but not least a certain syntactic and morphological structure. In contrast imagined voices are typically neither loud nor faint, neither sharp nor flat, and sometimes not even have a determinate syntactic or morphological structure.

Imagine a zebra and tell me how many stripes the zebra has. Unless you explicitly imagine the zebra to have a certain number of stripes, you will not be able to count the

stripes. The imagined zebra is indeterminate with respect to the number of stripes. This must not be interpreted as a situation of epistemic insufficiency. Consider for comparison the case of perception. Sometimes when you perceive a zebra, you are not aware how many stripes it has, either. Here, however, the zebra is determinate with respect to the number of its stripes. The number is determined by a fact of the matter. That you don't know the number has to do with your imperfect epistemic situation. In the case of imagination, in contrast, there is no fact of the matter that determines the number of stripes. We can even say that, *for you*, the number of stripes of the perceived zebra is determinate whereas, *for you*, the number of stripes of the imagined zebra is not. In case of the perceived zebra, it would make sense, for you, to approach the zebra or to use binoculars in order to find out how many stripes it has. If the number was not determinate from your point of view, this attempt should strike you as nonsense. In contrast, the idea of approaching the imagined zebra or viewing it through binoculars in order to count the number of stripes would be lunatic.

Can we explain the asymmetry between perception and hallucination on the one hand and imagination on the other? A suggestive hypothesis might be the following: In imagination properties are represented as properties as such, not as properties of some object. In perception and hallucination, to the contrary, objects or events are represented. Here, properties are represented only as instantiated by objects or events. Whenever they are represented as such, the subject does not have a perception or hallucination, but an imagination.

That the content of imagination is indeterminate has frequently been discussed especially when the question was at issue whether mental images are pictorial or propositional.⁷ Our hypothesis suggests that the content of imagination is neither pictorial nor propositional. It rather is constituted by a sequence of properties, something of the logical form $F_1 \& \dots \& F_n$, with F_1, \dots, F_n being properties. Like propositional representations imaginations are composed of concepts. But unlike propositional representations, the operation of predication is not employed. This is because, unlike perceptions and hallucinations, imaginations aren't cases of object or event consciousness. They are cases of pure property consciousness. Thus, imaginations do not assert anything; they don't even have a truth-value. Maybe this is something imaginations have in common with pictures or, even better, paintings. One can say "This painting resembles (or is true of) Mary" but one cannot say, "This painting is true". Paintings like imaginations are reminiscent of complex predicates that have not yet been predicated to an object. But unlike paintings, imaginations do not necessarily employ any topographical means of representation (recall the case of imagined voices.) To put it in terms of cognitive science: imaginations are conceptual property representations that have not been bound to an object representation. Using Fregean terminology, imaginations are unsaturated concepts (cf. Frege 1881/1986: p. 22).

We are aware that we aren't zombies because we have the capacity of imagination. When we imagine, we notice contents that can impossibly be real. This is because the

⁷ Cf. Armstrong (1968) and Tye (1991).

logical form of imagined contents is so different from the logical form of facts. When we imagine voices as in thinking, or when we imagine visual properties as in day-dreaming, we become aware of properties as not instantiated by either events or objects. This justifies us to infer that they can be nothing but contents of consciousness. We conclude that we have minds.

Objections and empirical defense

One may object that the argument for self-awareness by imagination presupposes a certain view of imagination that is just one possibility among others. Why should the scope of the imagination operation be devoid of bound variables or individual constants? The nature of imagination is an empirical issue, of course. One should, thus, be inclined to allow for abductive justification for the proposed view on imagination. Our theory predicts that the capacity of self-awareness should be positively correlated with the capacity of imagination. An impairment concerning the capacity of imagination should be positively correlated with an impairment concerning the capacity of self-awareness. Studies on autistic and schizophrenic patients seem to confirm these predictions. Autists have difficulties with imagination and also with understanding the idea of a mind. As Baron-Cohen (1995) observed, autists seem to be blind with respect to minds. They systematically make mistakes in attributing beliefs to others and to themselves. On the other hand, they also seem to have a reduced capacity of imagination. They do, e.g., not show pretense behavior. Schizophrenics, as a further example, often loose track of their thoughts (cf. Stephens & Graham, 2003). They suffer from so-called thought withdrawal, an experience that somehow one's thoughts have been taken away, removed as it were, from one's mind. This indicates that their awareness of themselves is disturbed. On the other hand, they tend to hallucinate when others only imagine. This is how the phenomenon of *gedankenlautwerden*, i.e., of hearing thoughts of one's own spoken aloud, is interpreted. Voices that ordinary people imagine when they silently talk to themselves become hallucinated voices in the schizophrenic case. In schizophrenia and autism, an impaired capacity of awareness of minds, thus, correlates with a disturbed capacity of imagination. The hypothesis that one knows of minds because one has the faculty of imagination explains this correlation. The question, how we know that we are conscious beings, thus, turns on empirical issues that, expected or not, have to do with our faculty of imagination.

References

- Armstrong, D. M. (1968). *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.
- Baron-Cohen, S. (1995). *Mindblindness*, Cambridge, MA: MIT Press.
- Carnap, R. (1928). *Der logische Aufbau der Welt*. Berlin: Weltkreis-Verlag.
- Chisholm, R. (1966). *Theory of Knowledge*. New York: Prentice Hall.

- Dretske, F. (1999). The Minds Awareness of Itself. In: *Perception, Knowledge and Belief*. Cambridge: Cambridge University Press: 158–177.
- (2003). How Do You Know You are Not a Zombie? In: Gertler, B. (ed.), *Privileged Access. Philosophical Accounts of Self-Knowledge*, Burlington: Ashgate Publishing: 1–14.
- Frege, G. (1881/1986). Funktion und Begriff. In: Patzig, G. (ed.), *Funktion, Begriff, Bedeutung*. Göttingen, 1986: Vandenhoeck & Ruprecht. (Original from 1881): 18–39.
- Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4: 31–52.
- Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The Theory of Event Coding (TEC): A Framework for Perception and Action Planning. *Behavioral and Brain Sciences*, 24: 849–937 .
- Moore, G. E. (1903). The Refutation of Idealism. *Mind*, 12: 433–53.
- Pianesi, F. and Varzi, A. C. (2000). Events and Event Talk. In: J. Higginbotham, F. Pianesi and A. C. Varzi (eds.), *Speaking of Events*. Oxford: OUP: 3–47.
- Sellars, W. (1953). Empiricism and the Philosophy of Mind. In: *Science, Perception and Reality*. London: Routledge & Kegan Paul.
- Stephens, G. L. and Graham, G. (2003). *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge: MIT Press.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6: 171–178.
- Tye, M. (1991). *The Imagery Debate*. Cambridge, MA: MIT Press.