

An introduction to human consciousness

Chris Paffen, 2019

Experimental Psychology & Helmholtz Institute, Utrecht University



Helmholtz  Instituut

<https://doi.org/10.23644/uu.7706573>

© Chris Paffen, 2019



CC BY-SA 4.0

Preface

This text discusses human consciousness, focusing on its nature: what is it, what is it not? I intend to give an overview of the philosophical positions that have been taken by several scholars, starting with Descartes. In philosophers terms, it deals with the ontology (what is the nature?) and teleology (what is its function, if any?) of consciousness. For this reason, I don't discuss contemporary (neuroscientific) theories on consciousness, such as Dennett's multiple drafts model (Dennett, 1991) or integrated information theory (Tononi, 2008). I would be very happy if readers would go on and learn about such theories while keeping in mind the philosophical considerations I outline in this text.

What is consciousness?

The first question to be answered is the obvious one: what do we mean with the term consciousness? I would say (guided by Chalmers, 1995) that we generally refer to one of two uses of the concept. The first use is related to uttering sentences like 'I wasn't aware of that', 'I did notice this', or 'at that time I was already asleep'. What these utterances have in common is that it involves reflecting on (one's own) thoughts, feelings and actions. It is also related to what Block (1995) labelled *access consciousness*, which he phrased as consciousness as being available 'for use in reasoning and rationally guiding speech and action'. Let's consider the following example to clarify this (also see Figure 1): Imagine being at a fruit stall with a friend. The fruit stall is very big and displays a lot of different fruits. Now your friend says: 'do you see those apples?'. The apples are right in front of you, but you are not aware of them. It takes



Figure 1 When you have searched a fruit stall for apples and become able to say that you have located them (as in A), Block would argue that you have access consciousness of the apples. When you are experiencing aspects of apples (such as its taste in B) he would argue that you are phenomenologically conscious of apples. Nagel further argued that you - as experiencer - are the only one knowing what it is like to experience this.

you a while to locate the apples, after which you reply ‘Yes, I see them, they are there.’ We can describe what happened here in terms of access consciousness: at first, you were not aware of the apples, even though the light reflected from them impinged on your retina. Thus, although your sensory organ (the eyes) sensed the apples, you had no access to the information sensed by the organ. After you had located the apples, you did become aware of the apples, and could reflect on you being aware of them: you had access to the information sensed by your eyes.

What then, is the other form of consciousness we can refer to? Let me say that this is the philosopher (of mind)’s use of the concept. The instances that the philosopher (and others) refer to being *phenomenologically* conscious of something. Let’s get right to the example I used above to illustrate this. Think of the apples in the fruit stall. Now imagine holding an apple in your hand, moving it to your mouth and taking a bite. How does the apple feel in your hand? What experiences do you get when taking a bite? Do you taste tartness or sweetness? Do you feel the juices in your mouth? Do you smell the sweetness? Imagining (or in fact actually taking a bite) invokes *phenomenal consciousness*, or, as Nagel (1974) labeled it: the ‘what’s it like’ aspect of consciousness. To discuss this kind of consciousness, Nagel used the metaphor of imagining to be a bat (Hence his article’s title ‘What’s it like to be a bat?’): we could imagine flying around in the dark, hanging upside down and even to eat bugs, but can we imagine what’s it like to be a bat? No we can’t, Nagel argues, and so it is the case with (phenomenal) consciousness: it’s subjective, private and based on introspection.

Can we study consciousness?

Why not?, you perhaps think, but philosophers thoughts and texts have complicated this for centuries, and the struggle continues. To begin looking at the struggle, let’s start with Descartes, as modern philosophy also often does. Among several groundbreaking arguments, Descartes is arguably best known for the phrase, ‘I think, therefore I am’, or as in the original text ‘Je pense, donc je suis’ (Descartes, 1637). Descartes came to this conclusion by using his famous ‘method of the doubt’. To paraphrase him, the argument would go something like this:

When seeking for the truth [whether things exist]:
I [Descartes] can doubt [the existence of] everything I perceive
If I doubt something, I cannot conclude that it exists
As a result, I have to refute its existence
I cannot, however, doubt that I have doubts
Doubting is a form of thinking
Something that thinks cannot, at the same time, be inexistent
I think, therefore I am

The highly influential conclusion Descartes reached, then, was that *thinking* exists, that is to say: thinking is real. Since thinking is an aspect of the mind, Descartes therefore provided proof that the mind exists. Before discussing the complication for studying consciousness, I first have to continue with Descartes, and discuss the proposition that ultimately lead to the *mind-body*

problem (of which I don't know who coined the term). After arriving to the conclusion that *thinking* exists, Descartes also concluded that matter exists. Again, my paraphrasing:

I can doubt the existence of my body

However, God¹ lets me have all kinds of sensations through my body (I feel pain, heat, emotions, et cetera)

God does not mislead me

I have to conclude that my body [also] exists

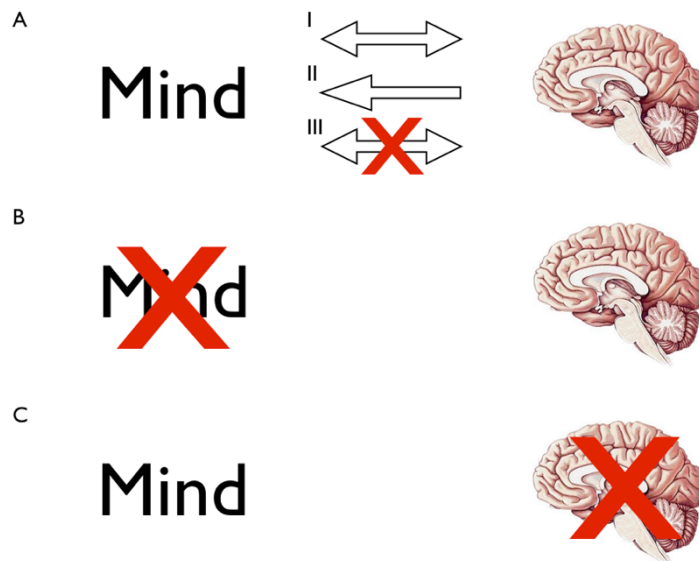


Figure 2: Three positions one could take in the 'mind-body' debate. A depicts the dualist's position: both mind and matter exist. Descartes argued there being a two-way interaction between the two (arrow AI). A dualist defending property dualism would say that the mind is a property of the brain. An epiphenomalist would also argue that the interaction is one-directional: the matter (in this case the brain) directing the mind (arrow AII). Posing that there is no interaction would make one a parallelist (AIII). B depicts the materialist's position: only matter (and thus a brain) exists. C makes one an idealist: only the mind exists.

Thus, Descartes argued that both thinking (an aspect of the *mind*²) and the body (a form of *matter*) exist. Where then, did all the trouble come from (the trouble in studying consciousness)? Arguably, this all started with Descartes point that 'extension in length, breadth and depth, constitutes the nature of corporeal substance; and thought constitutes the nature of the thinking substance'. Later known as *substance dualism*, this quote argues that mind and matter are made up of different *stuff*. Most scientists in the natural sciences won't have much problem with studying 'corporeal substance': matter and mechanisms within and between it are at the core of studying physics and chemistry, but how to study that other 'stuff'? What is the stuff that the mind (including consciousness) is made of? And, to directly turn to

¹ Descartes had already provided evidence for the existence of God. I won't discuss his argumentation here.

² Note that I will be talking about the mind for a while, and not specifically about consciousness. However, as being conscious is an aspect of the mind, issues pertaining to the mind also pertain to consciousness.

another problem which arose from Descartes' stance: if mind and matter are made of other substances, how could they possibly interact? Let's talk about the latter problem first.

(How) do mind and body interact?

As we saw above, Descartes argued that mind and body are made of different substances. Still, Descartes insisted that interaction between them was possible. Famously, he argued that the interaction occurred in the pineal gland, where 'animal spirits' interacted with the mind. Moreover, he claimed the interaction to be bidirectional: the mind controlling the body, but the body also influencing the mind. An example of the latter would be the following: imagine drinking a few glasses of wine. According to Descartes, this wine would influence your 'animal spirits', which in turn would affect your 'passions' leading to a change in your mind. In contemporary terms, we would translate this into something like: drinking wine changes the chemical composition of the blood, this in turn changes mechanisms (neurotransmitters and such) in the brain, which would make us feel tipsy (or drunk if we had too much of it). Another nice quote also illustrates Descartes' view on the interaction: 'If we see an animal approaching [...] the images in the brain form on the (pineal) gland, which, acting immediately upon the soul, causes it to see the shape of the animal'.

However, the issue of interaction between mind and body remained a problem: how can matter and non-matter interact? Philosophers have proposed several solutions for this problem, one introducing the concept of *property dualism*. According to this stance, mind and matter are not made of different substances: instead the mind is a property of matter. According to a modern view, we would say that the brain has different properties: it has a weight, size, is made of neurons, blood and other stuff. The mind, however, *also* is a property of the brain. Huxley's (T. H. Huxley, 1874) *epiphenomalism* is a clear example of property dualism: the mind (including consciousness) is the result of physical processes of the brain. As the brain causes our limbs to move, our mouth to speak and our skin to sense touch, it also causes aspects of the mind. Importantly, property dualists still insist that the mind is non-physical, implying that the natural sciences cannot be used to study it.

Before I move to two more radical solutions to the interaction problem, I'll discuss one more form of dualism, which, perhaps, solves the interaction problem. If the interaction is problematic, why not just conclude that there is no interaction? This is what – among others – was proposed by Malebranche (1688/1923): mind and matter simply don't interact: they act in parallel. While solving the interaction problem, this, of course, creates a new problem: how do mind and matter perform actions (seemingly) in concert? If I think of getting a drink from the refrigerator, this thought appears to guide my actions: I walk to the machine, open the door and get a can of soda. How then, are these thoughts and actions (of the body) orchestrated? Malebranche offers a solution here: God is the director, and orchestrates the timing between thoughts and actions. This, perhaps exotic, form of dualism is called *parallelism*³.

³ The view that mind and matter do not interact is called parallelism, the view that God orchestrates the interaction is called occasionalism.

Materialism

Above, I discussed one solution to the mind-body interaction problem (parallelism). However, other solutions are also possible. If the interaction is problematic and parallelism too exotic, we could simply deny the existence of one of the two. Let's start with denying the existence of the mind (as a substance or property different from matter), and take the position that there is only matter (*materialism* or *physicalism*). It's easy to see why this position has become very dominant in the natural sciences and in our time, also among the main disciplines studying aspects of the mind: psychologists, cognitive scientists and (cognitive) neuroscientists. From Newton onwards, materialistic theories and laws became highly apt to describe and predict phenomena in the natural world: why an apple falls from a tree with a certain speed profile, what causes thunder and why our nose starts running when we have a cold. These examples illustrate the power of mechanistic explanations and are part of the immense collection of scientific progress that made it possible to have all the virtues (and vices) of the modern world.

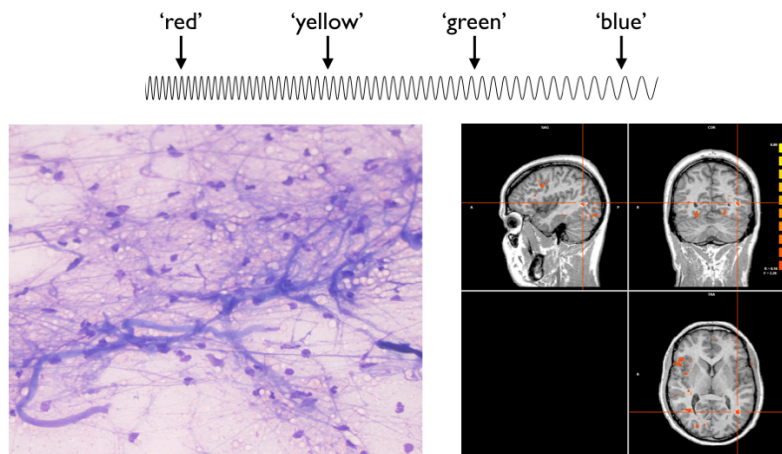


Figure 3 Materialists studying mental phenomena would only be interested in studying physical structures related to them. Reductionism would use mechanistic explanations such as: colour is related to the wavelength of light. This light can trigger reactions in nerve cells, organized in the brain (the picture depicts my brain). If we study those physical and biological substrates, can we explain mental phenomena?

As I said, *materialism* is currently (often unknowingly I suppose) dominant in the study of the mind. To quote a professor whose biological psychology class I once attended: 'the brain produces behavior as the kidney produces urine'. The latter phrase is an illustration of a particular omnipresent form of *materialism* which runs through (neuro-)biological psychology and other disciplines studying the mind: *reductionism* (Figure 3). According to this, a (mental) phenomenon is reduced to a mechanistic explanation. Of course, this has been very convenient in the natural sciences: I would prefer to describe thunder and lightning mechanistically (involving charged particles) instead of having to believe that some god is angry with us. Likewise, many scientists are happy to use reductionistic descriptions for mental phenomena. Yes, we can talk about what's it like to feel anxiety, but let's try to describe neural mechanisms (neurochemicals and brain-areas) which seem to cause it. Let's try to understand exactly how

a SSRI⁴ decreases fear in (a subset) of patients suffering from an anxiety disorder. Perhaps then, we don't have to keep doing intensive, time-consuming psychotherapies⁵. Let's take another example, which will lead us back to the philosopher's problem in studying consciousness, which I talked about earlier. Let's again think about the apples in the fruit stall, and imagine a green one. Let's say we are looking at the apple. Mechanistically, we could describe what happens as follows:

Some light source (say the sun) illuminates the apple. Part of this light is reflected back from the surface of the apple. The make-up of the surface of the apple is such that only part of the light is reflected back (light-waves within a certain frequency-range). These waves radiate towards our eyes and (partly) arrive at receptors in our retina. Some of these receptors are activated by the light-waves (those responding to a certain frequency range). I'm not too much of a neuroscientist, but I believe (part) of the process leading to activity in the receptor is the transformation of a molecule: from *trans-retinal* to *cis-retinal*. A cascade of chemical and electrical mechanisms follows: other cells in the retina become active, involving the traveling of currents through membranes, other chemicals within, and outside cells, ultimately leading to one of the central processes in the brain: the action potential. Action potentials start traveling from neuron to neuron, activating a new series of movements of currents and chemicals. The signals in the neurons travel toward the Lateral Geniculate Nucleus, are relayed towards primary visual cortex and further on. At some moment, signals arrive at a brain area called V4 whose neurons respond vigorously. The signal, however, keeps flowing: the sensation of receptors in the retina have activated other areas and processes. Additionally, the signals will not only travel in a feedforward way (as my description so far dealt with), but will be fed back (from 'higher' order areas, back to 'lower' areas, say primary visual cortex).

Now then, let's end the mechanistical description here. Notice anything? Read anything about perceiving the apple, or, for that matter, about the color green? This mechanistic description has led us to the problem: has the above explained *anything* about perceiving an apple to be green? In terms of *access* and *phenomenal consciousness*: have we learned anything about the labeling of an observer of an object to be green, or, more complicated even, about this observer *experiencing* the color green?

Back to studying consciousness

The above has introduced a fundamental problem in the study of consciousness: how can we explain this (or any) mental phenomenon using mechanistic explanations? Labelled the

⁴ Selective Serotonin Reuptake Inhibitor: a drug used in treating anxiety disorders.

⁵ I don't actually think there are a lot of psychologists who would want this. However, recently a Dutch insurance company (Menzis) decided to reimburse the treatment of depression based on results (Sondermeijer, 2018). Some magical drug would make ineffective therapies and the reduced reimbursement of them obsolete.

explanatory gap by Levine (1983), T.H. Huxley (1866) phrased it like this: “How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djinn when Aladdin rubbed his lamp in the story”. To face the problems more directly, we can again turn to Chalmers (1995), who dissociated between *easy* and *hard problems* of consciousness. Phenomena related to *access consciousness* belong to the easy problems. For example, not noticing something (an example I used at the beginning of this text) could be described in terms of attentional mechanisms: since the brain cannot process all incoming information, it selects part of the information (by some filter mechanism) for further processing. Information that is not selected won’t get (deeply) processed further. For this reason an observer cannot access the information and act on it (for example, uttering ‘I see an apple’ without lying). Explaining ‘not noticing something’ is already hard enough, but not so hard as what Chalmers (1995) referred to as the *hard problem* of consciousness: how can we describe *experience*, the *phenomenal*, or the ‘what’s it like’ aspect of consciousness? According to Nagel (1974), this phenomenal experience is reserved for the individual who has it, being the sole person being able to know what it is like. Observers other than the one having the experience, or science in general, can tell us nothing about it.

Idealism

Let us talk a little more about being an observer. Observing phenomena is arguably at the heart of modern science in general: wasn’t it Copernicus, later backed by Galilei, who challenged Aristotle’s geocentric⁶ model of the universe by observing that celestial bodies did not move according to the idea that the sun revolved around the earth? These observations lead Bacon (1620/1893) to argue that observations should be at the center of science: theories should be built on observations. But let us now realize a central aspect of making observations: they are made by observers who reflect on their experiences! How could Newton conclude that different colors emerge when a light beam hits a prism at a certain angle? Because he was *experiencing* the colors. In other words: Newton had *phenomenal* experiences of colors, and was able to *access* these experiences by writing them down. This insight lead Locke (1690) to radically propose that *all* knowledge was gathered by making observations: everything we will ever know is the result of experiences, and of the observers reflecting on these experiences (observers *accessing* their *phenomenal* experiences). By his stance, Locke started paving the way towards *idealism*: the idea that we can only be sure of the existence of our observations (or perceptions), and even have to reject the existence of matter.

Idealism is born by Locke’s discussion of primary and secondary qualities of observations: primary qualities are those that are fundamental to the external world, the qualities that Descartes used to describe matter as ‘extension in length, breadth and depth’. According to Locke, these qualities belonged to the physical world and were learned by us through observations. Secondary qualities, however, were qualities belonging only to the observer: touch, smell, color, taste et cetera. These qualities are not in the matter, even though we are used to saying that a pillow is soft, grass has a particular odor, tomatoes are red and

⁶ The earth at the center.

candy tastes sweet. According to Locke, all these qualities are in the observer, and not in the objects associated with them. To quote Newton (1704): “For the [light-]rays, to speak properly, are not coloured. In them there is nothing else than a certain power and disposition to stir up a sensation of this or that colour.”



Figure 4 Rembrandt's 'Jeremiah mourning the destruction of Jerusalem'. We perceive a lot of depth in the picture. But surely, we are not actually looking at a physical scene containing 'extension in length, breadth and depth': it's a picture of a painting and you're looking at it via a computer screen or piece of paper!

With Berkeley (1734) we finally arrive at idealism. In contrast to Locke, he claimed that primary qualities too, were limited to the realm of observations. For example, he observed that when we perceive the depth of objects in front of us (involving qualities of external space), we do not directly sense these qualities. Instead, we have learned about depths by experience. For example, previous encounters with objects has taught us that we have to reach further for an object that is partly occluded by another one. Being very much interested in optics, Berkeley also contemplated the seemingly paradoxical situation that light reflected from objects falling on our eyes is projected upside down on our retinas, although we do generally perceive it as such. To Berkeley this was not a problem: experience has taught us how to observe and approach these objects in ways that suit our behavioral goals. To Berkeley then, primary qualities were *also* within the exclusive domain of the observer. This notion implies that the world an observer perceives is *totally* the product of his or her mind. As he put it: "To be is to be perceived (or to perceive)". What about the physical world then? Does it actually exist in

some objective reality outside of observers? Again⁷, God comes to the rescue here: Berkeley proposed God to be the permanent perceiver: with God's eye, everything we only experience actually exists.

God's eye?

Berkeley's argument about the existence of the physical world by the grace of God sounds quite exotic to us now (to me at least). Be that as it may, physicist and computer scientist Koenderink (2014) lets us realize that this idea is actually, probably implicitly⁸, present in contemporary science. The important point brought forward by Locke and Berkeley is that we gather knowledge through observation, and that there is no other source for this. If we accept this viewpoint, and, for sake of the argument, let's do this right now, we have to accept that we cannot know anything about the world that we cannot observe⁹. The point Koenderink makes is that the observations we make are specific to the (nature of the) observer. In a talk I attended, Koenderink explained this point by talking about leeches. Leeches are quite limited in what they can sense and perceive, they have no vision, smell or taste, and probably no sense for the passage of time. About the only thing they do is that they hang in a tree for years, until, at some point, they fall down – on a living creature if they're lucky – and suck blood. If we accept that the world is defined by the nature of the observer, what kind of world do leeches live in? What kind of world do bats live in? Let's do another thought experiment: let's say that some life form, other than those known on earth, evolved somewhere in the universe, with other senses than those known to us now: what kind of world would they live in? In von Uexküll's (1909) words, all creatures live in their own 'Umwelt'. In Koenderink's terms, most scientists believe that the progress of scientific knowledge is not limited by the observer: we can know everything there is to know about the (physical) world, because we can make *objective* (i.e. not limited by the observer) observations: the world as it really *is* (in God's eye view). But what about the aspects of the world for which we don't have senses? A way to talk about this is to make a distinction between unknowns - that what we can observe, but haven't observed yet, and unknowables - that what we cannot observe, and therefore can never know about. We can approach unknowns with our scientific tools (allowing us to *observe*), but how could we possibly study the unknowables (the stuff we cannot observe)? The latter also implicates that we should be careful about what could threaten our existence on earth: we can make predictions on how climate change or the sun's increase in size might someday harm us, but what about unknowable threats? Predictions based on what we can observe are helpful, but only to the point until 'shit happens' (Koenderink's wording). To end this section, Koenderink teaches us that whenever a scientist (psychologist, physicist, biologist or whatever kind) makes claims about the world, the correct response is: whose world are you talking about?

⁷ Like in the argumentation of Descartes and Malebranche.

⁸ I do not envision scientists believe in a God observing the world.

⁹ T.H. Huxley's grandson Aldous wrote it like this: "That which [...] is called "this world" is the universe of reduced awareness, expressed, and, as it were, petrified by language" (Huxley, 1954).

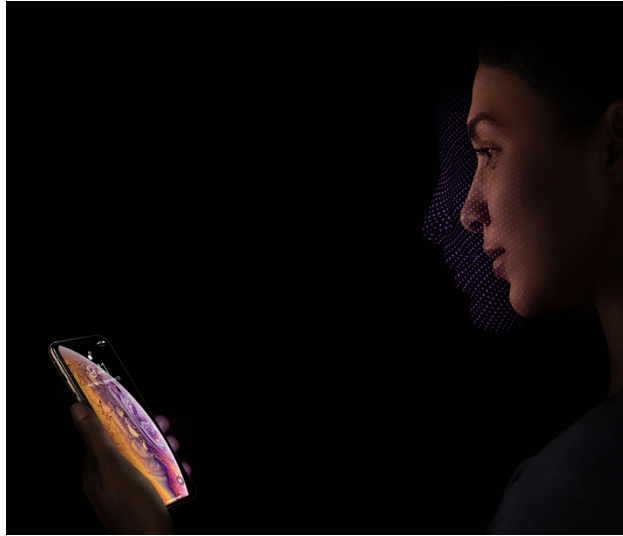


Figure 5 An iPhone can be unlocked by directing it towards a face, making its face recognition functional. But could it be conscious of the face?

Functionalism

The above overview of dualistic and monoïstic interpretations of mind and matter might lead one to believe that one has to pick one of three positions: of being dualist, materialist or idealist. In addition, it seems as though serious scientists would opt to choose *materialism*, in order to avoid the difficulty of studying the mind (as being non-material). Another option is to just ignore talking about the mind, and asking: what does the mind do (as opposed to what is it made of)? This is essentially what *functionalism* does. This notion has been brought to us by modern day cognitive science in the wake of the birth of the computer. In the conceptualization of functionalists a mind can produce behaviors, as – so goes the analogy – computers can. To use a modern day example: an iPhone can recognize faces, as we humans can. The important concept here is *information processing*. Minds and computers process information, in order to produce behavior. As Neisser (2014), who stood at the cradle of cognitive psychology, put it: “A program is [...] a recipe for selecting, storing, recovering, combining, outputting, and generally manipulating *information*. [...] this means that programs have much in common with theories of cognition.”. Information processing is one of three fundamental steps. First, there is input (say the projection of some optical information on a retina or camera). Next, computations are performed - information is processed - on this input (say algorithms working on the optical structure that seek for facial features). The final step is the output of the computation. In our face example, the output could be the unlocking of the phone (the algorithms of the iPhone identifying the optical structure as the owner of the phone) in case of the iPhone, and the greet ‘Hé Chris, how are you doing!’, in case of the human.

A central consideration in functionalism is that it is of no relevance to talk about the nature of the mind; it’s all about its function. In addition, it doesn’t really matter what *stuff* performs the computations: in our example, both the iPhone and the human observer ‘recognize’ a face. The latter observation is of course a very optimistic notion when considering the possibilities of artificial intelligence. If the function of computations is the only thing that

matters, and if it doesn't matter what kind of body performs the computations, every aspect of human information processing (one can also say all human cognitive competences) can in principle be implemented in a machine. If consciousness is such a competence (can it be, in functionalism?) then we should be able to create it.

Philosophical zombies

Even though functionalism declares the nature of the mind irrelevant, some do challenge the existence of at least one aspect of the mind: consciousness. Most famously, Dennett is ardent at declaring its inexistence¹⁰. Let me give one example of the line of arguments he puts forward, involving philosophical zombies. Let's start with imagining two different kinds of creatures, living next to each other. Importantly, the creatures are basically identical, except for the fact that only one kind possesses phenomenal consciousness (these we call humans, the ones without it zombies). Both creatures appear to behave the same (for someone observing them): they talk, they move, and go about their daily business. However, in his reasoning, (phenomenal) consciousness has no adaptive value: the creatures that have it are not more 'fit' than the creatures without it. According to Dennett's reasoning then, natural selection (through producing offspring in the creatures) would favor the zombies, as their makeup is supposedly simpler. As he writes: "had some time traveling hyperengineer inserted a *robot-replicator*¹¹ into the milieu, and if its prowess was equal or better than the prowess of its natural grown competition, its descendants might now be among us - might even *be* us!" (Dennett, 1991) Note that Dennett puts forward a functionalist argument here: the consciousness of the human creatures has no adaptive value (no functionality), and thereby loses the struggle for existence against the simpler creatures without it. But what's it like to be a philosophical zombie?

Meaningless computation

The above thought experiment involving zombies illustrates the functionalist standpoint: if some creature can produce functional properties like recognizing faces, it is of no relevance that the property might be accompanied with an experience (being conscious of) of that property. There is, however, another issue that comes up when thinking about functionalism, which is *intentionality*. The concept is used differently by different scholars, but here I will discuss it as referring to mental states 'being directed at things'. The 'being directed at things' was introduced by Brentano who used it to describe the observation that when I think about a chair, this thought is connected with actual chairs; the thought is directed at a chair. As Brentano (1874) stated it: "Obviously there is no act of thinking without an object that is thought, nor a desire without a thing that is desired"¹². According to Brentano, only mental phenomena can have intentionality (being directed towards an object). Could philosophical zombies have intentionality? Or, in more contemporary terms, could a computer have intentionality?

¹⁰ Although his arguments mostly apply to the phenomenal aspect of consciousness (see also Block).

¹¹ This is akin one of the creatures in my example: an agent possessing no phenomenal consciousness.

¹² The original German text of Brentano: "Offenbar gebe es kein Denken ohne ein Object, das gedacht, kein Begehren ohne einen Gegenstand, der begehrt werde." Brentano, 1874, pp 116.

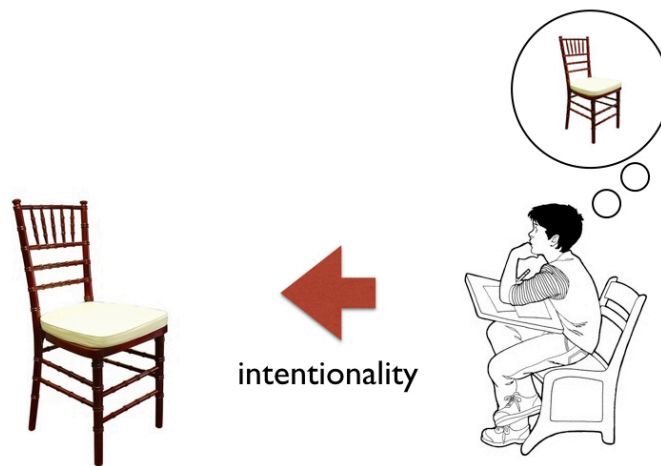


Figure 5 Thoughts (and mental states) are said to be: they are directed at things.

Before turning to this question, let's consider another issue that comes up when thinking about functionalism, which is *meaning*. I will introduce an example which will lead us to a famous thought experiment of Searle (1980). Let's suppose we ask a person or a computer a question. Actually, you can try this yourself. The question is: 'What is color?' If I ask you this question, you will probably come up with an answer, since you *understood* the question; it meant something to you. You might return that color is a quality of objects in the outside world, or, if you are convinced by certain points in this text, a quality that arises in the perception of an observer. What now if we ask this question to a computer. This you can do yourself, you can type in Google 'What is color'. When you press enter you will get answers! The real question here is: did your computer (or the algorithm Google uses to provide you with answers) *understand* this question? We have arrived at Searle's Chinese Room experiment. Paraphrased, this experiment goes something like this:

Envision a typical Dutch person. This person is most certainly capable of understanding Dutch language (written, spoken) and can also communicate in this language. Probably, this person is also capable of understanding some other languages, like English, German or French. The person probably does not, however, understand Chinese. Let's envision such a person, being incapable of understanding spoken Chinese, and, more importantly, of the written characters the language uses. To this person, Chinese characters are meaningless scribbles, that don't mean anything to the person. This is the first step.

In the second step, the person is asked to pick a card from a container (container Q) containing many cards. When a card is picked, it contains a list of Chinese symbols and an instruction (in Dutch!!) of which cards to pick from a second and third container (container S and container W). Each card in the second container contains more Chinese symbols. The card in the third container has another set of instructions (also written in Dutch!). The instruction is something like: 'write down the symbols from container W in the order: Card 4023, Card 3222, Card 321, Card 123, Card 23'.

This is of course very abstract, but let's now realize the following: the cards in container Q contain questions written in Chinese, for example: 'can machines think?'. The cards in

container W contain Chinese words and expressions. In our example, they were 'it' (card 23), 'doubt' (card 123), 'can' (card 321), 'they' (card 3222) and 'if' (card 4023). The instructions on the card from container S contained of course the order in which to put the cards from container W. What is the conclusion from this? The Dutch person has given a sensible answer, which he didn't understand, to a question which he didn't understand either! The latter point is exactly what Searle is arguing: we can device a program (resembling the instructions mentioned above) that can link certain input (cards from container Q) to certain output (cards in a certain order from container W). This program produces functional, and perhaps sensible output. It however lacks one fundamental aspect: *meaning*. A machine understands as much as the Dutch person in the Chinese room experiment: nothing. When I ask you the question 'What is color?', you both understand the question as well as the answer you provide me with; when you ask Google, neither is the case. Needless to say that the card-picking of the Dutch person also doesn't involve intentionality. Do they have mental states that can be directed at things? As Searle put it: '...intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output.'

Consciousness as giving meaning

I am nearing the end of my short essay on consciousness. I intentionally built up the story in the last couple of paragraphs to get to the point Searle made: without consciousness, there seems to be no place for both intentionality and meaning. Here I again arrive at Koenderink (2011). His view on consciousness centers on it providing *meaning* to the observations we make. Since Koenderink is as much a scientist studying vision as he is physicist and computer scientist, I will borrow an argument involving the human visual system. Let's return to an example I used earlier: the observation of an apple. As laid out above, we can describe how the eyes and the brain respond to the apple; from the light rays to the retina, from the retina to the lateral genicular nucleus (LGN), from the LGN to the cortex. We can even envision that all this processing leads some neurons to signal that there was light with a certain wavelength, and that there was a spatial structure arising from the contrasts between the 'green' light and the background. However, as we learned from Searle, just performing computations doesn't lead to *knowing*, not does it lead to *meaning*. Koenderink labels the computations of the neurons 'pointer readings' (outcomes of some computation): they are meaningless. Then what is consciousness? According to Koenderink, consciousness involves 'probing the structure' (of say the apple): investigating it. When the probing meets something we have encountered before, meaning is created: we are conscious of the apple. Let's take a moment to dissect this, what is Koenderink trying to say? The rays that enter the eye clearly lead to a sensed optical structure: neurons point toward qualities and locations in the structure. At this stage, a zombie or AI device could do the job: an iPhone can provide the optical structure of a face. But what's next? In Koenderink's terms, the optical structure is investigated: based on what we know and on past experiences, we provide hypotheses about the structure: is it an orange, or a melon perhaps? What is the evidence telling us? Hmm, there is something in the nature of the light rays, I have encountered this before, they have a characteristic grass also has. In addition, I ate, touched, smelled this thing before. Hypotheses (oranges, melons) are eliminated: a new best

guess: an apple! Surely, a zombie or AI device could do this too: match some structure against a database, and provide a best guess. But then, we become aware of the apple, with all the encounters we had with it, we understand it is an apple, we experience it, we are conscious of it.

Acknowledgements

I would like to thank Andrea van Doorn and Stefan van der Stigchel for useful comments on this manuscript.

References

- Bacon, L. (1620/1893). *Novum organum (True suggestions for the interpretation of nature)*. New York: George Routledge and Sons.
- Berkeley, G. (1734). *The Principles of Human Knowledge*. London: Jacob Tonson.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227–287.
- Brentano, F. (1874). *Psychology vom empirischen Standpunkte*. Leipzig: Duncker & Humblot.
- Chalmers, D. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2, 200–219.
- Dennett, D. C. (1991). *Consciousness explained*. New York: Back Bay Books.
- Descartes, R. (1637). *Discours de la méthode*. Leiden: Jan Maire.
- Huxley, A. (1954). *The doors of perception*. New York: Harper & Brothers.
- Huxley, T. H. (1866). *Lessons in elementary physiology*. London: MacMillan and Co.
- Huxley, T. H. (1874). On the Hypothesis That Animals Are Automata, and Its History. *The Fortnightly Review*, 16, 555–580.
- Koenderink, J. (2011). Vision as a user interface. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging XVI*, Vol. 7865, 786504. <http://doi.org/10.1117/12.881671>
- Koenderink, J. (2014). The All Seeing Eye? *Perception*, 43(1), 1–6. <http://doi.org/10.1007/978-3-662-24819-5>
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64, 354–361.
- Locke, J. (1690). *An Essay Concerning Human Understanding Book II: Ideas*. London: Thomas Bassett.
- Malebranche, N. (1688/1923). *Dialogues on metaphysics and on religion*. (M. Ginsdale, Trans.). London: George Allen & Unwin Ltd.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83, 1–9.
- Neisser, U. (2014). *Cognitive Psychology*. New York: Psychology Press.
- Newton, I. (1704). *Opticks, or, a treatise of the reflections, refractions and colours of light. Also, two treatises of the species and magnitude of curvilinear figures*. London: Sam Smith.
- Searle, J. (1980). Minds, brains, and computers. *Behavioral and Brain Sciences*, 3, 1–8.
- Sondermeijer, V. (2018, August 13). Menzis gaat behandeling voor depressie anders vergoeden. *NRC Handelsblad*.
- Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3), 216–242. <http://doi.org/10.2307/25470707>
- Uexküll, von, J. (1909). *Umwelt und Innenwelt der Tiere*. Berlin: Julius Springer.

