Neuronal agnosticism and constructive empiricism

Ju Young Lee
Graduate Training Centre of Neuroscience
University of Tübingen

1. Introduction

In this paper, I will argue that constructive empiricism is a viable way of understanding neuroscience. Constructive empiricism establishes the rationality of agnosticism about the claims that scientific theories make on unobservables. It was originally defended by Ban van Fraassen in his book “The Scientific Image” published in 1980. His formulation of constructive empiricism is as follows:

Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate [van Fraassen. 1980, page 12].

In order to unpack what these claims amount to, I will first explain what empirical adequacy is. Van Fraassen suggests that a theory is empirically adequate when the claims it makes about the observables are true [van Fraassen. 1980, page 18]. This means that an essential part of constructive empiricism is drawing a line between observables and unobservables. Van Fraassen argues that drawing this line allows us to see scientific theories as aiming to account for observable processes by postulating unobservable processes [van Fraassen. 1980, page 3]. Although it is quite reasonable to suppose that there is an intuitive line that divides observables from unobservables, the way that van Fraassen draws the distinction between them is rather provocative. He only considers unaided acts of perception by members of an epistemic community as acts of observation [van Fraassen. 2001, page 15]. Therefore, any activity that involves using tools, such as using a microscope or telescope, would not count as an act of observation [van Fraassen. 2001, page 154].

2. Instruments: Windows or Engines

The typical metaphor that is used when talking about scientific instruments is that they function as windows to the invisible level of nature. This metaphor lends itself naturally to thinking of instruments as revealing features of nature that are beyond direct perception. But van Fraassen thinks that using instruments does not count as observation. What, then, are instruments doing if not aiding observation?

Van Fraassen proposes a new metaphor for thinking about instruments. Instead of thinking about instruments as “windows”, we should think of them as “engines” that create new observable phenomena. [van Fraassen. 2001, page 154-155]. For example, consider an experimental setting where a dish filled with pond water is placed under an optical microscope. If someone looking through a microscope spots a paramecium, she will think that the optical microscope magnified the pond water and revealed a micro-organism that she could not see with unaided vision. But this only makes sense on the metaphor that instruments function as windows. Once we start thinking about them as engines of creation, it becomes clear that what instruments do is create images. These images (like the image of the paramecium) are images of the observables they are pointed at (in this case, the pond water). The key point is that in this new metaphor, the image of the paramecium does not reveal the microstructure of the pond
water. The image of the paramecium only exists under this measurement setup and is dependent on the instrument. We should instead consider the image of a paramecium to be a new observable phenomenon created by the optical microscope [van Fraassen. 2008, page 441]. Similarly, any instrument that seems to “reveal” what is not observable should be regarded as an “engine” that produces new observable phenomena of the targeted object.

3. Constructive Empiricism for biological neuroscience

This view may seem rather plausible when applied to the domain of physics, where the relevant instruments include cloud chambers or particle colliders. Nonetheless, we might worry that casting doubt on microscopes will be much more problematic for biologists. For example, in neuroscience, the fundamental entity is the neuron. But a neuron cannot be observed with unaided perception. Thus, according to constructive empiricism, the neuron is an unobservable entity. This may seem problematic since the idea that neuronal networks are composed of synaptic connections between discrete neurons is a fundamental part of how neuroscientists study the brain. However, I think that this concern is overstated. To show why, we should look at the history of neuroscience.

There used to be a long debate on the exact microstructure of the brain. On one side, Cajal proposed a “neuronal doctrine” which postulated that a brain is comprised of discrete units called neurons [Cajal. 1888]. On the other side, Gerlach and Golgi proposed a “reticular theory” [Gerlach. 1871]. The reticular theory postulated the structure of the brain to be a continuous single network. Experimental evidences such as optical microscopy of stained nervous system were more comprehensible under neuronal doctrine than reticular theory [Nansen. 1887]. However, Golgi continued to defend reticular theory by modifying the theory and emphasizing the anatomy of a diffuse nerve network [Golgi. 1906]. Ultimately, the debate came to an end with the development of the electron microscope. Electron microscopes revealed the existence of synapses between neurons and led to a definite finding supporting the neuronal doctrine [López-Muñoz and Álamo. 2006].

Nevertheless, following constructive empiricism, instruments are engines not windows. This means that images created by a microscope are new observable phenomena that are not necessarily magnifications of the targeted slides. If images from an electron microscope do not contain information concerning the microstructure of a brain sample, then the existence of neurons, let alone synapses, cannot be taken for granted. Therefore, neuronal doctrine cannot be vindicated nor can reticular theory be refuted.

When most of the scientific knowledge concerning neuroscience supposes synaptic connectivity, constructive empiricism sounds as if it is denying the entirety of neuroscience. However, I would like to emphasize that refutation of reticular theory is not necessary in order to carry out research based on neuronal doctrine. Researchers can build hypotheses and design experiments based on a theory even if the theory has not been vindicated or even if a rival theory exist. If an interpretation of the data made assuming the neuronal doctrine is reasonable, then it is possible to use neuronal doctrine but remain agnostic about the truth of the theory’s claims about unobservables. Thus, constructive empiricism sees the rationality of designing experiments and analyzing data based on neuronal doctrine. The important difference is that constructive empiricism allows agnosticism of the postulated processes that are ultimately inaccessible with unaided perception.

4. Constructive empiricism for neuroimaging research
When studying the human brain, non-invasive neuroimaging methods are frequently used. Examples of such methods include magnetic resonance imaging (MRI) and electroencephalography (EEG). By using different sequences of the MRI scanner, one can acquire a structural image of the brain, as well as a functional image, referred to as functional MRI (fMRI). According to constructive empiricism, EEG and the MRI scanner should be considered as engines of creation instead of windows to the brain. One might instantly think that such stance creates a big discrepancy in how a neuroscientist interprets the images produced by the instruments. However, I would like to argue that neuroscientists with a constructive empiricism stance can interpret the signals recorded by the instruments just like any other neuroscientist. In the following sections, I will first present the standard theories of the instruments and then propose an alternative stance for constructive empiricists for both methodologies.

4.1 Functional magnetic resonance imaging

fMRI is interpreted as containing spatial information about which neurons are activated in the brain. The theories involved in fMRI are quite complex, involving theories of brain physiology as well as the physics of MRI itself. I will first go through the postulated unobservable processes relevant to the brain physiology. Then, I will discuss the physics of the MRI scanner which also involves many unobservable processes. Afterwards, I will explain how these theories of brain physiology and MRI physics are combined to support the claim that the fMRI signal provides spatial information concerning activated brain regions. Lastly, I will demonstrate how fMRI can be interpreted validly as brain signal by constructive empiricists.

The most important physiological entity for explaining the fMRI signal is a protein called hemoglobin. Hemoglobin is an oxygen transporter, which is loaded with oxygen molecules in the lungs and transports oxygen to sites throughout the body. Hemoglobin is embedded in red blood cells which travel through blood vessels. As oxygenated hemoglobin travels through vessels, the oxygen eventually detaches from hemoglobin and diffuses through the vessel barrier to reach nearby cells.

Cells in the brain can be categorized into two groups, neurons and glial cells. The exact glia-neuron ratio is still under discussion, however the most current works predict the ratio to be around 1:1 [Bartheld et al. 2016]. When neurons in a particular brain region are activated, the neighboring vessels dilate to bring in more red blood cells carrying oxygenated hemoglobin. The physiological linkage between neuronal activation and vessel dilation is called neurovascular coupling. Currently, there are several competing theories that try to account for the causal relationship between the two processes. Some researchers claim that vessels dilate due to direct innervation of activated neurons [Bonvento et al. 1991]. Others argue that vessels dilate in response to chemical signal from the activated neurons themselves [Iadecola et al. 1997] or in response to chemical signal released by glial cells near the activated neurons [Parri and Crunelli. 2003]. Nevertheless, it is commonly agreed that activation of neurons leads to dilation of surrounding vessels, which in turn increases the concentration of oxygenated hemoglobin, washing away the deoxygenated hemoglobin in the vessels.

MRI itself is based on the theory of nuclear magnetic resonance. Here, I will explain the physics of MRI using a gradient echo (GE) sequence as the example. The MRI scanner creates a strong magnetic field using a huge magnet. All hydrogen nuclei that are within this magnetic field become magnetized. The GE sequence starts with the transmission of a radiofrequency (RF) pulse. The RF pulse is then absorbed by the magnetized hydrogen nuclei. After the pulse, the hydrogen nuclei re-emit the absorbed energy as RF wave according to their relaxation time. The relaxation time of the hydrogen nuclei is dependent on the homogeneity of the
local magnetic field. Hydrogen nuclei in an inhomogeneous magnetic field have faster relaxation times compared with the hydrogen nuclei in a homogeneous magnetic field. The receiver coil in an MRI scanner measures the RF energy emitted from the hydrogen nuclei after a predetermined echo time. Since hydrogen nuclei with a fast relaxation time will have already emitted most of its RF energy, the receiver coil picks up RF energy emitted by hydrogen nuclei with a slow relaxation time, which are located near a homogeneous magnetic field. The signal recorded by the receiver coil is called the blood-oxygen level dependent (BOLD) signal and this signal is used to infer physiological processes occurring within the brain [Weishaupt et al. 2008].

A GE sequence can be used to indirectly measure brain activity because deoxygenated hemoglobin and oxygenated hemoglobin have different magnetic features. In particular, deoxygenated hemoglobin causes inhomogeneity in the local magnetic field due to its paramagnetic characteristics. In contrast, oxygenated hemoglobin is diamagnetic and does not cause inhomogeneity in the local magnetic field. Since the inhomogeneity of the magnetic field causes nearby hydrogen to have a faster relaxation time, much of their RF energy will be released before the receiver starts recording. The magnetic property of hemoglobin and vessel dilation together result in the hemodynamic response of BOLD signal as shown in Fig. 1.

**Figure 1. Hemodynamic response when neurons are activated.** [Kornak et al. 2011]

*a) Initial dip:* Prior to blood vessel dilation, there is a high quantity of deoxygenated hemoglobin due to oxygen uptake by the activated neurons. Increase of deoxygenated hemoglobin in the blood vessel generates strong inhomogeneity in the local magnetic field.

*b) Primary response:* Dilation of vessel and overcompensation of blood flow wash away the deoxygenated hemoglobin. Low quantity of deoxygenated hemoglobin in the blood vessel generates very weak inhomogeneity in the local magnetic field.

When neurons are activated, they take up oxygen from nearby vessels. Therefore, the concentration of deoxygenated hemoglobin in the vessel is high immediately after neuronal activation. A high quantity of deoxygenated hemoglobin leads to a strong inhomogeneity in the magnetic field and eventually to a low BOLD signal. This is reflected in the early stage of the hemodynamic response as an initial dip. However, it is well known that nearby blood vessels will dilate in response to increased neural activity and this increased blood flow will wash away the deoxygenated hemoglobin, causing an overcompensation for the earlier dip in the BOLD signal. This overcompensation rises well beyond baseline because the blood vessel dilation efficiently washes away all deoxygenated hemoglobin, resulting in a very low quantity of deoxygenated hemoglobin. This concentration ends up being much lower than the baseline, resulting in a greatly increased BOLD signal. It is this overcompensation that fMRI measures. Therefore, a high BOLD signal can be interpreted as activation of neurons located in that region following the inference below:
Step 1: Increased magnetic resonance (MR) signal in a specific region implies low concentration of deoxygenated hemoglobin
Step 2: Low concentration of deoxygenated hemoglobin implies dilation of local vessels
Step 3: Dilation of local vessels implies activation of the neighboring neurons

This is the standard theory of how fMRI functions. Considering this, the fMRI signal is seemingly an indirect method for measuring neuronal activity. However, I would like to argue that scientists can use fMRI while staying agnostic about the standard theory. In order to present the argument, I need to begin by listing the observable entities in the experimental setting that survive van Fraassen’s strict definition. The observable entities that can be accessed by direct perception are the operation of the MRI machine, the target object that is placed inside the scanner, the observable activity of the target object, and the recorded MR signals. Examples of observable activity of the target object include vascular dilation in the brain and task performance of the subject. There are certain regularities between these observables.

Regularity 1: When the MRI machine is on and the object is placed outside of the scanner, the recorded MR signal is zero
Regularity 2: When the MRI machine is on and the object is placed inside the scanner, the recorded signal is positive
Regularity 3: When the MRI machine is off, the recorded signal is zero regardless of the object’s position
Regularity 4: Differences in observable activities of the object being measured tend to correlate with differences in the recorded signal

Following van Fraassen’s metaphor, the MRI scanner is an example of an engine that creates MR signals as new observable phenomena. From the listed regularities, one can build the theory of MRI scanner that says the following:

An operating MRI scanner produces an MR signal, which is informative about the targeted object, through a set of unobservable processes.

No matter how complex the unobservable processes are, the above theory is empirically adequate if it is true in its claims about these observable regularities. For brain scientists, the targeted object is normally a subject’s head. The subject’s head has a sophisticated anatomy which includes the skin, muscles, scalp, and many layers of meninges and blood vessels, before we reach what we call the brain. Based on the postulated unobservable processes underlying the interaction between the MRI scanner and the subject’s head, she can invent an MR signal processing pipeline to extract information about the brain. She can choose to believe the theories of brain physiology explained above and interpret the extracted brain data accordingly. However, if she is a constructive empiricist, then she is also allowed to stay agnostic about the claims the theory makes about the unobservables. However, this agnosticism does not prevent her from being able to interpret the data according to the widely accepted theories of neurophysiology. The benefit of taking an agnostic stance towards the theories underlying fMRI is that the presence of the gaps in our understanding of the unobservable processes involved in fMRI, such as neurovascular coupling, become less problematic. The presence of such gaps in our understanding of the relevant unobservable mechanisms does not, on its own, threaten the empirical adequacy of the standard theory of fMRI.

4.2 Electroencephalography
An EEG signal can be interpreted as brain data by a constructive empiricist in a similar way. Theories of EEG are also complicated and multi-layered. However, you will see that the reasoning of constructive empiricism is applied in the same way as fMRI. I will begin by presenting the standard theory of how EEG works and then discuss the way the constructive empiricist understands EEG.

The most important physiological process in the theory of EEG is the ionic current between the intracellular and extracellular spaces of neurons. Molecular biology has postulated that the neuronal membrane has a complex structure that involves many different kinds of ionic channels. An activated neuron propagates its action potential via the influx and efflux of ions. The ionic flow between the neuronal membrane inevitably produces electrical current in the intracellular as well as extracellular space. The electrical current travels through conductive substances, which enables the extracellular current to reach the scalp where the EEG electrodes are placed. According to Ohm’s law, current can be calculated when the voltage and resistance are known. Since the EEG electrodes measure voltage fluctuation, one can infer electrical events occurring within the head by calculating the electrical current.

However, the standard theory of EEG has issues. Different tissues within the brain have different impedance depending on its type. The skull also has a different impedance. Since the electrical current produced by neuronal activity has to travel through different tissues and the skull, attenuation of the original signal is unavoidable. The problem of calculating how the original signal is transferred to the electrode is called the “forward problem”. The head models that scientists use to solve this problem have gradually evolved. However, the in vivo conductivity of different tissues is still unknown and the influence of changes in parameters of the head models needs further investigation [Hallez et al. 2007]. Additionally, glial cells such as astrocytes also have several types of ion channels embedded in their membrane. The flow of ions through these channels inevitably result in an electrical current in the extracellular space [Parri et al. 2010]. Therefore, the voltage recorded by EEG electrodes is a combination of neuronal and glial electrophysiology. However, exactly how much glial electrophysiology contributes to the EEG signal is unclear.

Constructive empiricism can transform these gaps that are seemingly problematic into trivial matters. As in the previous section, I can build a theory of EEG by considering the regularities between the observables. The observables involved in the experimental set up are the EEG machine, the target object upon which EEG electrodes are placed, the observable activity of the target object, and the recorded EEG signals. Examples of the observable activity of the target object include eye blink movements and task performance. The regularities between these observables are the following:

Regularity 1: When the EEG machine is on and the electrodes are not attached to the object, the recorded EEG signal is zero
Regularity 2: When the EEG machine is on and the electrodes are attached to the object, the recorded signal is positive
Regularity 3: When the EEG machine is off, the recorded signal is zero regardless of the object’s position
Regularity 4: Differences in the observable activities of the object being measured tend to correlate with differences in the recorded signal

For the constructive empiricist, the EEG machine can be viewed as an engine that creates EEG signals as new observable phenomena. From the listed regularities, one can build a theory of the EEG machine that says the following:
An operating EEG machine produces EEG signals, which is informative about the targeted object, through a set of unobservable processes.

The above theory is empirically adequate because it is true in its claims about the observables. Additionally, the gaps in theory become part of the unobservable processes and do not influence the empirical adequacy of the theory of EEG. A researcher may build theories on the origin of the recorded EEG signals and how it propagates to the scalp. Based on the postulated unobservable processes underlying the interaction between the EEG machine and subject’s head, she may invent an EEG signal processing pipeline to extract information about the brain. The key difference of being a constructive empiricist is that she is also allowed to stay agnostic about the claims the theory makes about unobservables. Despite her agnosticism, she can still interpret the data according to the widely accepted theories of neurophysiology and electrodynamics.

4.3 Benefit of constructive empiricism in neuroimaging research

Some researchers prefer to use fMRI over EEG due to the availability of a highly detailed account of the neurophysiological processes underlying the hemodynamic response pattern of the BOLD signal. In contrast, the features of EEG recordings such as the N100 and P300 peaks are more difficult to explain based on neurophysiology. We do not have a mechanistic understanding of the electrophysiological processes generating these regularities in the EEG signal. Nevertheless, some people prefer EEG over fMRI because there are fewer steps of inference between the signal itself and the neuronal activity. If one accepts the fact that neuronal activity causes some electrical current, then based on Ohm’s law, one can directly conclude that EEG recordings are informative about neuronal activity.

The benefit of being a constructive empiricist is that the researcher can choose to treat fMRI and EEG as equally informative, so long as both instruments are empirically adequate. If one methodology is not more empirically adequate than the other, then there is no reason to believe that one is more reflective of brain activity than the other. Since it seems that our theories of fMRI and EEG are empirically adequate, we can treat them as equals. One disadvantage of constructive empiricism is that it is possible to postulate two different unobservable processes to explain the images generated by two different instruments aimed at the same target. At first sight, this may appear to prevent scientific progress. However, constructive empiricism does not endorse relativism [Bueno. 2003]. All constructive empiricism does is provide the option of staying agnostic regarding which of these unobservable processes are literally true. In this sense, constructive empiricism might open up more opportunities for science by allowing a more pluralistic approach and preventing us from getting stuck in dogmatism.

5. More pragmatic benefits for brain researchers

5.1 Benefit of becoming a constructive empiricist as a brain scientist

A fundamental belief underlying all the work done by brain scientists is that mental processes, like emotion and cognition, are tied to processes in the brain. However, explaining exactly how the mind and brain are related is a difficult problem known as the “mind-body problem”. In contemporary philosophy of mind, there are three main options for solving this problem: reductive materialism, non-reductive materialism, and dualism. Reductive materialism assumes that the relationship between mind and body is something like an identity relation, non-
reductive materialism assumes that it is more like a supervenience relationship, and dualism assumes that it is something more like a causal relationship. Since it is very difficult to imagine how we might gather scientific evidence that supports one of these theories over the rest, it would be beneficial to be able to remain agnostic about the exact relationship between the mind and the body.

Happily, a constructive empiricist can choose to remain agnostic with respect to these candidates for the mind-body relationship. Nevertheless, she does have to believe that the two observables, the brain and the mind, are related somehow. This is because of several regularities observed between the two. The famous cases of Phineas Gage [Larner and Leach. 2002] and Henry Molaison [Corkin et al. 1997] are representative examples of how brain lesions can lead to changes in a subject's cognition. Additionally, the effects of deep brain stimulation on Parkinson's patients and patients with major depression [Mayberg et al. 2005] provide another regularity that establishes that there must be some kind of relationship between the mind and body. Finally, research performed on brain-machine interfaces has enabled paralyzed patients to move robots [Hochberg et al. 2012]. This gives additional evidence that there must be some sort of relationship between the two.

5.2 Healthier attitude for working scientists

For a working scientist, positioning herself as a constructive empiricist can relieve her from the burden of searching for truth. Scientists often worry that the evidence they have might be insufficient to infer the reality of nature, a problem called the underdetermination of theory by evidence. Let's imagine a scenario where a memory researcher builds a conceptual model of how memory works and formulates a mathematical equation based on the model. She conducts several experiments measuring an electrical parameter (for example, the local field potential) in the hippocampus and memory task performance and finds out that her model successfully explains the relationship between the two measurements. Despite the high accuracy rate, she may still have doubts concerning her model of memory and think that the model provided successful predictions because it was partially true. She might even doubt the model being partially true and worry that the success was merely a coincidence. Since the physicality of memory is unknown, the model cannot be validated for its correspondence to the truth.

Constructive empiricism can save scientists of such despair by providing an alternative stance. By aiming for the validation of the empirical adequacy of her theory instead of validation of its truth, she can save herself from the worries of her theory not being true. The effort can then be directed at something possible to accomplish, achieving the empirical adequacy of her theory. In this example, the theory of memory would be empirically adequate because it is true about its claims on the regularity between the two observable measurements. The truth of the model is not so important anymore and the successful explanation of the observable processes by the model is definitely something to celebrate.

6. Conclusion

In this paper, I demonstrated that constructive empiricism is a viable way of understanding different fields of neuroscience. In the first and second section, I introduced the key concepts of constructive empiricism. Constructive empiricism is a philosophical position that defines the aim of science as providing empirically adequate theories. The empirical adequacy of a theory is achieved if the claims the theory make about the observables are correct. According to van Fraassen, the distinguishing line between observables and unobservables is drawn where the limit of unaided perception lies. In contrast to common views on instruments,
which is to consider them as windows to the invisible part of nature, van Fraassen suggests that we should consider them as engines of creation.

In the third and fourth sections, I explained how neuroscientists could take a constructive empiricism stance and interpret data acquired from the head as neuronal data. Neuroscience can still operate normally even when instruments are viewed as engines that create new observations. As long as there are regularities between the targeted object, the instrument, and the measured signal, the signals can be interpreted by supposing unobservable processes and structures, such as the neuronal doctrine. The important difference when taking a constructive empiricism stance is that the neuroscientist is allowed to stay agnostic about any of the claims the theory makes about unobservables. In the fifth section, I argued that there are some potential pragmatic benefits available to brain researchers that accept constructive empiricism. First, they are allowed to leave aside the hard problem of the mind-body relationship. Second, by aiming for empirical adequacy instead of truth, one can worry less about the problem of underdetermination.

Finally, I would like to emphasize that constructive empiricists do not consider it irrational to believe the claims that scientific theories make about unobservables. Van Fraassen does not say that agnosticism is more rational than believing the claims of the unobservables. His point is that agnosticism about unobservables is a valid epistemological stance as well:

I respect the rationality of those who prefer to have those supererogatory beliefs, my arguments are meant to show the rationality of those who’d forego them. [van Fraassen. 2001, page 168]

7. Acknowledgement

I would like to thank Wesley Sauret for his overall supervision, especially concerning philosophical reasoning, and Johannes Stezle for his validation of scientific statements.

8. Reference


Nansen, F., 1887. The structure and combination of the histological elements of the central nervous system. J. Grieg.


