Philosophia (2017) 45:371–385
DOI 10.1007/s11406-016-9750-7

On Extreme versus Moderate Methodological Naturalism

Feng Ye[1]

Received: 23 June 2016 / Revised: 14 August 2016 / Accepted: 17 August 2016 /
Published online: 25 August 2016

Abstract In a recent debate, Rosenberg claims that only the methods of natural science can deliver genuine knowledge, while Williamson rejects Rosenberg's extreme methodological naturalism and insists that we have genuine philosophical and humanistic knowledge not achievable by hard-scientific methods alone. This paper responds to the debate. I will argue that physicalism, together with contemporary neurocognitive and evolutionary knowledge, implies that some of our intuitions and mental simulations used in the humanities and philosophy are justified methods for achieving knowledge but are *practically ir*replaceable with hard-scientific methods. That is, extreme methodological naturalism is in conflict with physicalism. The argument also shows that some moderate version of methodological naturalism can be consistent with physicalism. Physicalism is the strong version of *ontological* naturalism and is supposed to be accepted by strongly committed naturalists like Rosenberg. Therefore, to be self-consistent, these naturalists should adopt physicalism (as ontological naturalism) plus a *moderate* version of methodological naturalism, rather than Rosenberg's extreme version.

Keywords Methodological naturalism · Ontological naturalism · Intuition · Mental simulation

1 Introduction

Rosenberg and Williamson have a debate on methodological naturalism recently (Rosenberg 2014a, b; Williamson 2014a, b). Rosenberg claims that *only* the methods of the hard sciences (i.e., the experimental methods of natural science) can deliver genuine knowledge and that traditional humanistic and philosophical beliefs can become genuine knowledge *only if* they can be reduced to hard-scientific beliefs and

✉ Feng Ye
fengye63@gmail.com; http://cnu–cn.academia.edu/FengYe

[1] Department of Philosophy, Capital Normal University, 83 The Western 3rd Ring Road North, Beijing 100089, People's Republic of China

achieved by hard-scientific methods. This is a rather extreme version of methodological naturalism. Williamson argues that Rosenberg's extreme methodological naturalism is untenable, because we do have humanistic and philosophical knowledge *not* achievable by hard-scientific methods alone. This paper will respond to the debate. I will argue that physicalism, which is the strong version of *ontological* naturalism,[1] together with our neurocognitive and evolutionary knowledge, actually implies that some human cognitive methods are justified methods for achieving knowledge but are *practically ir*replaceable with hard-scientific methods. This means that extreme methodological naturalism is in conflict with strong ontological naturalism. The argument will also show that some moderate version of methodological naturalism can be compatible with physicalism. Therefore, to be self-consistent, strong naturalists like Rosenberg should adopt a *moderate* (rather than extreme) version of methodological naturalism, together with physicalism as strong ontological naturalism.

Therefore, this paper will take physicalism and contemporary neurocognitive and evolutionary knowledge as premises, and at the same time it will *not* assume anything violating the spirit of physicalism when justifying cognitive methods or demonstrating their practical irreplaceability. Section 2 will first give a general idea about how a cognitive method can be justifiable but practically irreplaceable under physicalism. Here I will consider a special class of human cognitive abilities, complex pattern recognition, which allow humans to recognize, directly and effortlessly, some very complex patterns in human environments (including their own activities). It will be argued that some of these complex pattern recognition abilities are justified methods for achieving knowledge but are *practically ir*replaceable with ordinary perceptual abilities and other hard-scientific methods. I will show that some common intuitions and mental simulations used by philosophers and other humanistic researchers involve instances of complex pattern recognition as their critical steps. Therefore, it can potentially happen that some of these intuitions and mental simulations are also justifiable but *practically ir*replaceable.

Then, Section 3 argues more specifically that some of our basic psychological intuitions and mental simulations for recognizing and predicting human emotions, desires, beliefs and behavioral tendencies are justifiable but practically irreplaceable methods. These intuitions and mental simulations are our basic methods for achieving knowledge in the humanities. Moreover, as an example from philosophy, Section 4 will argue that our epistemological intuition for judging Gettier cases is also justifiable but practically irreplaceable. These examples show that Rosenberg's extreme methodological naturalism is in conflict with physicalism.

The argument to this point already shows that the existence of justifiable but irreplaceable methods is compatible with physicalism. Section 5 will show further that physicalism can explain why hard-scientific methods and other irreplaceable methods have their idiosyncratic features, and why in some situations hard-scientific methods have advantages, but in some other situations we have to rely on those practically irreplaceable methods. That is, physicalism can provide a coherent account of *both* hard-scientific methods *and* other irreplaceable methods.

[1] See Papineau (2007) and Stoljar (2009) for a general introduction to naturalism and physicalism.

Therefore, in sum, a strong but coherent naturalistic worldview should consist of physicalism (as strong ontological naturalism) plus a *moderate* version of methodological naturalism, not Rosenberg's extreme version.

There have been heated debates on philosophical intuitions and other related methodological issues in recent years.[2] The approach in this paper differs from other approaches in that its starting point is physicalism (and can be any radical version of it), but a little surprisingly, its conclusion on intuitions and methodological issues turns out to be closer to traditional and more conservative commonsense views.

Here are a few more clarifications before I proceed.

First, to be more definite, I will assume that hard-scientific methods include ordinary perceptual abilities for recognizing simple physical objects and states (which are needed for observing scientific instruments), as well as deduction, induction, abduction and other experimental and theory-building methods in the natural sciences.

Second, I will not try to give an exact definition of the moderate version of methodological naturalism compatible with physicalism. This will require a general characterization of *all* methods justifiable under physicalism, which is beyond the scope of this paper. It suffices to note that these methods should include the justifiable but irreplaceable methods discussed in this paper, and therefore the version can't be Rosenberg's extreme methodological naturalism.

Third, here I take physicalism as a premise, but I am not suggesting that physicalism is a *strictly necessary* premise for concluding that there are justifiable but irreplaceable methods. If you do not like naturalism or physicalism, very likely, you already believe that there are many such methods. However, strong naturalists like Rosenberg may claim that those methods in your mind are in fact *un*justifiable because of presupposing something violating the spirit of naturalism. Therefore, my strategy is to demonstrate both the justifiability and irreducibility of some methods *strictly under physicalism*. On the other side, the argument here will explicitly assume that a human pattern recognition ability is realized by a neural mechanism. Therefore, physicalism *is* needed as a premise here.

Fourth, I have no intention to claim that knowledge obtained by intuitions and mental simulations is a priori. The emphasis here is merely that some intuitions and mental simulations do deliver genuine knowledge but are practically irreplaceable with hard-scientific methods.

Finally, I always emphasize *practical* irreplaceability here. Perhaps we can agree that a *fictional*, *ideal* intelligent agent with *unlimited* computational power can, *in principle*, replace every acceptable cognitive method by hard-scientific methods. However, methodological issues are about what methods we can *practically* use and are entitled to use. Methods that *only* a fictional, ideal intelligent agent can use have nothing to do with methodological strategies for us humans. Moreover, to reject extreme methodological naturalism, we only need to show that some methods *are* justifiable but are *very likely* practically irreplaceable, for as long as they are so, insisting on replacing them with hard-scientific methods is not a good methodological strategy.

[2] See Pust (2012); Goldman (2010); Haug (2014); Symons (2008) and Vaidya (2010) for some surveys and discussions.

2 Complex Pattern Recognition

I will take physicalism as a premise from now on, and I will be careful not to assume anything violating the spirit of physicalism. Physicalism implies that human cognitive processes are completely physical. 'Physical' here (as usual) is understood in a broad sense, including biological and neurological entities and properties. Therefore, physicalism actually implies that human cognitive processes are neural processes within human brains and physical interactions between brains and their environments. Then, a human cognitive method is a pattern of neural activities controlled by a neural mechanism in a brain. In particular, each pattern recognition ability is realized by a neural mechanism in the brain. In recent decades, scientists and engineers are fervently simulating human pattern recognition abilities by using artificial neural networks. I won't assume here that these artificial neural networks correctly model how human neural mechanisms work. It suffices to recognize that human neural mechanisms *can* perform pattern recognition.

Ordinary perceptual abilities are relatively simple pattern recognition abilities, but here I am more concerned with complex pattern recognition, which is part and parcel of many intuitions and mental simulations (or thought experiments) considered by philosophers. For instance, consider our psychological intuition for judging that someone is angry. We can recognize a complex pattern among human facial, vocal, gestural and other behavioral features, namely, the pattern of angry appearances. When judging that a person is angry, we first recognize that his or her facial, vocal, gestural and other behavioral appearances instantiate that pattern, and then we infer that he or she is angry. The first step here is an instance of pattern recognition. Note that the second step is actually an inference to the best explanation, or abduction. That is, from the observed angry appearances we infer that a person has the inner emotion of anger. Abduction is considered a standard scientific method. Therefore, what is special in this psychological intuition for recognizing anger is exercising a pattern recognition ability. This ability is much more complex than the ordinary perceptual abilities for recognizing ordinary physical objects such as scientific instruments, because it requires synthesizing appearance features of very different kinds, namely, facial, vocal, gestural and many other features, and it concerns some dynamic features. Here I will just consider it an instance of *complex* pattern recognition.

Note that it does not matter in case someone insists that this pattern recognition is *just a kind of* perception. The point is, as I will argue later, psychological intuitions like this allow us to achieve genuine knowledge in the humanities *and* they are *not practically* replaceable with the ordinary perceptual abilities for recognizing ordinary physical objects (together with other hard-scientific methods). Continuity between complex pattern recognition and perception is exactly what forces us to take a more liberal stance regarding methodological naturalism.

Complex pattern recognition is also the basis for many mental simulations. For instance, consider a mental simulation of a rock rolling down a hill. This can be used to draw some general conclusions about the movements of an object in similar situations in the real world (Williamson 2007, p.142). This involves *imagining* an event, namely, a rock rolling down a hill. However, note that we do not imagine the event *arbitrarily* in the mental simulation. For instance, we do not imagine that the rock tangos down the hill. We want the imaginary event to be sufficiently similar to the *real* events of objects

rolling down, so that we can draw sensible predictions, based on the imaginary event, about what will really happen for a rock in similar situations in the *real* world. This requires recognizing a pattern among the *real* events of objects rolling down and recognizing that our *imaginary* event instantiates that pattern. This is the critical step in the mental simulation. Now, this is recognizing some complex pattern among *events*, which have a temporal dimension. It is again unlike recognizing static physical objects based on their static appearances. It is thus an instance of complex pattern recognition and is different from perception in the ordinary sense.

In other words, at least in some cases, intuitions and mental simulations are nothing mysterious. They are not 'seeing by the mind's eyes', for instance. Their essential steps are just recognizing more complex patterns in human environments by more complex neural mechanisms (developed in evolution, presumably).

In some cases, we can verify the reliability of a pattern recognition ability among normal human subjects. We can do this by performing psychological experiments on human subjects. Moreover, sometimes we can do this without using our own very same pattern recognition ability. That is, our verification can be non-circular. Furthermore, sometimes we can have neurocognitive and evolutionary evidences for the existence of a neural mechanism realizing a pattern recognition ability, and especially, sometimes we can have an evolutionary explanation for the reliability of that ability.

Now, suppose that we can have all these for a pattern recognition ability. Then it seems that we can accept that ability as a justified method for achieving knowledge. That is certainly true if you hold reliabilism regarding epistemic justification, since we can verify the reliability of that method. The common objection to reliabilism as a theory of epistemic justification comes from the concern that there can be cases of unexplainable reliability. This will not happen here since we have assumed that there is a neurocognitive and evolutionary explanation about the existence and reliability of that pattern recognition ability. Such a neurocognitive and evolutionary explanation also shows that that pattern recognition ability can be considered an epistemic virtue of us endowed by evolution. Therefore, virtue epistemologists can also accept this ability as epistemically justified. Moreover, such a neurocognitive and evolutionary explanation also shows that exercising that pattern recognition ability coheres with our overall physicalistic views about humans and human cognitive methods. That is, that ability is also justifiable even if you prefer a coherent theory of epistemic justification.

Complex pattern recognition abilities differ from ordinary perceptual abilities only in that the former involve more complex neural mechanisms for recognizing more complex patterns in human environments. They can naturally be seen as a sort of 'more complex and possibly less reliable perception'. In case we *can* verify that such an ability is sufficiently reliable *and* we can have naturalistic explanations for its existence and reliability, then, as naturalists, we *have* to admit that it is a justified method for achieving knowledge.

Here I must clarify that, to accept a complex pattern recognition ability as a justified method, we do not have to wait until we have a *detailed* neurological theory about how exactly a neural mechanism realizes that method and how it is reliable. It will suffice as long as there are sufficient evidences, by ordinary scientific standards, that there *is* such a neural mechanism. For instance, we are still far from knowing the details of our neural mechanisms realizing visual perceptual abilities, but all neural scientists agree that we already have sufficient evidences for the *existence* of such neural mechanisms.

Sometimes we can have sufficient scientific evidences for the existence of something with some qualities and functions without knowing all details of that thing.

In the above I have listed some conditions under which we can accept a complex pattern recognition ability as a justified method for achieving knowledge. That is, we can accept it, in case we can verify its reliability, have scientific evidences for the existence of a neural mechanism realizing it, and have an evolutionary explanation for its reliability. Now consider the issue of replaceability with hard-scientific methods.

It seems that, under some other conditions, a brain cannot *practically* replace its exercises of a neural mechanism realizing a *complex* pattern recognition ability by exercises of ordinary perceptual abilities and other hard-scientific methods. Here are three such conditions: (1) an exercise of the mechanism is an integrated cognitive step not consciously analyzable by the brain itself into a combination of other explicit cognitive steps; (2) the neural mechanism consists of billions of neurons working together and is a result of hundreds of millions of years of evolution, which makes it *practically* impossible to simulate, by explicit linguistic computations and inferences, how those neurons in the neural mechanism work together to recognize a pattern; (3) the pattern recognized by the neural mechanism is also extremely complex, and it is *practically* impossible (or at least, so far we are very far from being able) to describe all details of the pattern *explicitly* in languages, by referring to ordinary perceptual features of the components of the pattern. When a neural mechanism and the pattern it recognizes meet these conditions, *practically*, there appears to be no way for the brain to stop using that mechanism and replace it with an explicit, conscious combination of a series of cognitive steps using *other* cognitive methods, including ordinary perceptual abilities and other hard-scientific methods.

Here is a concrete example for illustrating these. Consider our human facial recognition ability. Neural scientists believe that we have a special neural module (besides the common visual modules) for recognizing human faces. We can verify the reliability of this facial recognition ability among people, and we can do this without using our own facial recognition modules. For instance, we can show several photos of the same person to experimental subjects to test whether the subjects can recognize that they look like the same face. Here we can use other means to make sure that they *are* photos of the same face, without using our own facial recognition modules. That is, this verification is non-circular. Moreover, facial recognition is critical for the survival of human families and societies. That is, there is an evolutionary explanation for the existence and reliability of our neural mechanisms for facial recognition. Therefore, this facial recognition ability meets the conditions listed above for a justified method for achieving knowledge.

On the other side, it also satisfies the three conditions listed above for practical irreplaceability. It appears that no brain can possibly stop using that special neural module for recognizing faces and replace it with other cognitive methods such as explicit, rule-based, linguistic computations and inferences on perceived basic visual features of a face. That is *practically* impossible simply because of the huge amount of *explicit, conscious, linguistic* computations and inferences (not unconscious neural computations) required. (We do use unconscious neural computations to perform such visual pattern recognition tasks.) Our uses of our facial recognition modules seem practically irreplaceable.

Note that using other means to decide whether two photos are from the same face in order to test the reliability of a subject's facial recognition ability is *not* a replacement for our facial recognition ability, for the ability is for deciding whether two photos *look like* the same face, and it does not really matter whether they *are* from the same face. We just use those other means to verify the reliability of people's facial recognition module. That is, sometimes we can verify the reliability of a method by using other methods without being able to replace the former by the latter.

I will not discuss here whether this facial recognition ability is just *a kind of* perception. The point is that our complex pattern recognition abilities used in some of our intuitions and mental simulations in the humanities and philosophy are not essentially different from this facial recognition ability in terms of their justifiability and practical irreplaceability. If that is true, then we have to admit that some methods used in the humanities and philosophy are justifiable but not practically replaceable with hard-scientific methods, which supposedly include only *ordinary* perceptual abilities for recognizing *ordinary* physical objects such as scientific instruments. This already refutes extreme methodological naturalism.

In sum, some complex pattern recognition abilities are justifiable, because they are essentially just 'a sort of more complex perception' and are endowed to us by evolution, but they are not replaceable with ordinary perceptual abilities and other hard-scientific methods, because, as results of a long history of evolution, they are realized by extremely complex neural mechanisms, and *practically*, we cannot fully analyze those neural mechanisms or the patterns they recognize so as to replicate their functions by using *other* methods.

3 Some Methods in the Humanities

Now, let's examine some justifiable but irreplaceable methods in the humanities and philosophy. First, consider our psychological intuitions for recognizing emotions, desires, beliefs and behavioral tendencies in another person. For instance, consider our psychological intuition for recognizing anger in other people. As I have explained in the last section, this intuition involves an instance of complex pattern recognition for recognizing the pattern of angry appearances.

This is a pattern determined by a hidden cause, namely, the emotion of anger. We can verify the reliability and stability of this pattern recognition ability among normal, competent people. Here, we can use independent means to ascertain that a person *is* angry when we use that person to test a subject's such ability. For instance, based on experiences about our own emotions and behaviors, we can associate a particular cause of anger with our own anger and further associate it with a particular behavioral sign of anger. We believe that other normal people are just like us. This allows us to monitor or even manipulate a person's anger, by observing or manipulating that cause of anger and by observing that behavioral sign of anger. Note that this does not require us to use our own ability for recognizing the *general* pattern of angry appearances. That is, here we can make sure that a person is angry by observing a *particular* cause and sign. Then, we can test an experimental subject's ability for recognizing the *general* pattern of angry appearances by masking that particular cause of anger from the subject and letting the subject observe only angry appearances. We can do this without using our

own same ability for recognizing the general pattern of angry appearances. This then can provide non-circular evidences for the reliability of people's ability for recognizing the general pattern of angry appearances.

Note that here one can either assume that anger is an inner state or take a more behavioristic stance in interpreting anger, which does not seem to affect our conclusion on reliability. Nor does it matter even if you hold the extremely physicalistic view that the mental state of anger is *identical with* a hidden neural state. In that case you will agree that we can rather reliably identify that hidden neural state by its effects on one's appearances, although we cannot recognize that neural state directly.

On the other side, we notice that very young children already have this pattern recognition ability. The ability seems mostly innate. We believe that since we are social animals, evolution must have endowed us with such innate abilities for recognizing psychological states in others rather reliably, which is a precondition for us to form a sophisticated society. Therefore, we have reasons to believe that there *is* a neural mechanism realizing this pattern recognition ability and that there *is* an evolutionary explanation for its reliability, although we do not yet know the details. Consequently, this pattern recognition ability meets the conditions for justifiability listed in the last section, and we can accept it and the psychological intuition based on it as justified methods for achieving knowledge.

Now, similar to the case of facial recognition ability, the neural mechanism realizing this ability may involve billions of neurons working together in a single, not consciously analyzable cognitive step. This neural mechanism responds to detected features in a person's face, voices, gestures and other behaviors and recognizes a complex pattern there. We do not expect that a human brain can replace a use of this mechanism with a combination of other explicit cognitive steps, for instance, by recognizing some more basic facial, vocal and gestural features and then performing *explicit, linguistic* computations and inferences to decide whether those features constitute an angry appearance. That is, our uses of this mechanism seem *practically* irreplaceable.

This is not to deny that we can have a *partial* reductive description about angry appearances. A good art teacher, for instance, can summarize some of the facial features that make a face looks like an angry face. The point is: a *complete* description of such facial, vocal, gestural and all other features seems *practically* impossible, and moreover, it is not true that having such a *complete* description is a *precondition* for our uses of this intuitive cognitive ability to be epistemologically justified. This is already sufficient to conclude that it is *not* a good methodological strategy to insist on replacing our uses of this psychological intuition with hard-scientific methods.

Our other psychological intuitions for recognizing emotions, desires, beliefs and behavioral tendencies in other people seem to be similar to this one. They also involve irreplaceable instances of complex (sometimes very complex) pattern recognition. We can accept many of these intuitions but have to admit that they are very likely not practically replaceable by hard-scientific methods.

Our psychological cognitive abilities are not limited to recognizing *instant* emotional states, desires, beliefs and behavioral tendencies. We can also predict the developments of a person's emotional and other mental states. These can be done by mental simulations. For instance, we observe many real instances of emotional changes, including their causes and their behavioral signs, and we keep them in memory. Then, when we observe a person's current emotional state and try to predict his or

her emotional changes in a near future, we imagine an episode of emotional changes starting from that observed current emotional state, and we try to make that imaginary episode instantiate the pattern characterizing those *real* instances of emotional changes kept in memory. Then, we can use that imaginary episode to predict that person's emotional changes. As I have explained in the last section, the critical step in such a mental simulation is recognizing a complex pattern among the *real* events of people's emotional changes in the real world and recognizing that the imaginary episode instantiates that pattern.

We similarly have reasons to believe that there is a neural mechanism realizing this ability of mental simulation although we do not know the neurological details. Moreover, complex pattern recognition abilities used in such mental simulations are even more complex than those in intuitions for recognizing instant emotional states. It is even less possible that we could replace such a mental simulation with conscious, explicit linguistic computations and inferences on observed basic facial, vocal, gestural and other features. That is, it seems quite obvious that such mental simulations are practically irreplaceable.

On the other side, the reliability of such abilities (i.e., mental simulations) can vary greatly depending on many factors, such as contextual conditions, one's expertise (innate intelligence as well as previous trainings) in understanding human emotions, and the temporal scale of a prediction. For instance, a short-term prediction can be quite reliable, but a long-term prediction is typically much less reliable. Again, the reliability of such a prediction is testable non-circularly, as long as we can use our own psychological intuitions for recognizing *instant* emotional states.

Then, consider an exercise of such a mental simulation by some person in a situation. If we have evidences to believe that that person is (comparatively) very good at understanding people's emotions and the reliability of previous similar exercises of that ability by her in similar situations is high, then we can admit that this exercise of the ability does deliver knowledge. Otherwise, if we have evidences to believe that someone is quite incompetent at understanding other's emotions, then we can deny that his prediction is knowledge, even if the prediction turns out true. The important point is: when we deny that an exercise of this ability in a situation delivers knowledge, it is not because we hold extreme methodological naturalism and accept only hard-scientific methods; it is because we have positive evidences to believe that the exercise of that ability cannot be sufficiently reliable in the situation (because of the complexity of the matters involved, or because of the incompetence of the person using that ability, or both). Our judgment is not much different from the commonsense judgments.

The same applies to our psychological and social-psychological abilities for predicting even more complex matters, for instance, for predicting mutual influences between people's emotions, desires and beliefs in a society, or even predicting the long-term social developments of people's psychological characters and behavioral tendencies. An exercise of such an ability is perhaps much more complex than simple mental simulations. It may involve imagining a rather complex episode with many people involved, and then recognizing that the imaginary episode is sufficiently similar to real events in the real world in relevant aspects. The reliability of our abilities for such predictions can be quite low. We can admit that our exercises of these abilities in typical cases at most deliver only good opinions or informed guesses, not genuine knowledge. However, once again, this is not because we hold extreme methodological naturalism

and accept only hard-scientific methods. It is because we have commonsense and/or scientific evidences to believe that one cannot really make any fairly reliable prediction on such extremely complex and versatile matters.

In sum, contrary to Rosenberg's extreme methodological naturalism, there *are* justifiable but practically irreplaceable intuitions and mental simulations in the humanities. On the other side, if we do think that a method in the humanities can deliver *only* good opinions or guesses, not genuine knowledge, it is not because we hold extreme methodological naturalism. It is because we have positive reasons to believe that the method can't be sufficiently reliable because of the complexity and versatility of the matters involved.

4 The Epistemological Intuition about Gettier Cases

Now, let's examine an example from philosophy. Consider our epistemological intuition about Gettier cases. First, note that we actually recognize *two* different patterns here, at least after we are aware of Gettier cases. The first is the pattern JTB, namely, 'justified, true, belief'; the second is the pattern of genuine knowledge, which I will denote as K here. Before Gettier, most philosophers failed to distinguish these two patterns. After Gettier, some philosophers claim that K should be analyzed as JTB + X with some very complex X, but recently, influenced by Williamson (2000), many philosophers deny this and believe that K is actually more primitive and more easily recognizable by humans than JTB, JTB + X (for any potential X), or even their component J, B, or X is. Either way, most contemporary epistemologists seem to agree that JTB and K are two distinct patterns.

To see that there *is* this pattern K, we only need to refer to the fact that most *well-trained* contemporary epistemologists are rather consistent in judging whether an example is an instance of genuine knowledge. In particular, when they discuss Gettier cases, they seem to agree among themselves about whether a case is a case of genuine knowledge. Here, the fact that some people *without* sufficient philosophical trainings disagree on those examples does not affect our conclusion. It is a common phenomenon that trainings *greatly* enhance people's pattern recognition abilities.

Moreover, we can even concede the possibility that the pattern K has *no objective, hidden* essence. When we recognize the appearances of tigers, for instance, we do recognize a pattern with a hidden essence, namely, tiger's genes, because appearances are determined by genes for animals. In contrast, for a token to be a token of the letter 'a' is *just* for it to be recognized so by human brains with sufficient linguistic trainings. The pattern 'tokens of the letter "a"' is therefore defined by referring to a group of human brains, not by a hidden essence of those tokens themselves. Now, under physicalism, a group of human brains is a group of physical systems. A pattern so defined *is* real, as long as those physical systems are quite stable and mutually consistent in the relevant aspect for defining that pattern. Similarly, it is possible that K is a pattern defined by referring to the brains of well-trained contemporary epistemologists'. That will still make it a real pattern, as long as those epistemologists *are* stable and mutually consistent in recognizing that pattern among human cognitive activities. Under physicalism, a conventional pattern *is* a real pattern, because the so-called 'conventions' are themselves physical facts, namely, patterns of activities of human brains as physical systems.

Furthermore, we can also concede the possibility that someday we will discover that the pattern JTB is somehow more relevant and useful for our *scientific* explanations about various aspects of human cognitive activities than the more easily recognizable pattern K is. Even if that happens, it is still true that the pattern K is real.

In other words, when an epistemologist says, 'Gettier cases are not cases of genuine knowledge', she can potentially be making two distinct assertions:

(1) There is a pattern K, which we intuitively call 'knowledge' but is distinct from the pattern JTB, and Gettier cases instantiate JTB but not K;
(2) The pattern K can play a more significant role in our *scientific* explanations about human cognitive activities in the future than the pattern JTB can, and therefore only K really deserves the name 'genuine knowledge'.

She may make both assertions based on her intuitions, but apparently (2) is much less certain than (1) is. For example, here is a question related to (2) whose answer is so far quite unclear. We agree that, in normal situations, true beliefs are more likely to be useful (for the agent holding the beliefs) than false beliefs are. We can perhaps agree further that a true belief is even more likely to be useful when it is justified, rather than unjustified. Now, is it true that a belief is *even more* likely to be useful when it is a K, a case of so-called 'genuine knowledge' recognized by current epistemologists, rather than merely a JTB? The answer is quite unclear, and I am not aware of any in-depth investigation on this question. If the answer is 'yes', then it supports (2). Otherwise, if the answer is 'no', then it supports the idea that the pattern JTB is perhaps more relevant for explaining aspects of human cognitive activities than the so-called 'genuine knowledge' is, even if the latter is more easily recognizable *intuitively*. The fact that a pattern is *intuitively* more primitive, more easily recognizable and less analyzable is not a guarantee that it carves Nature deeper at her joints and is thus more relevant and useful in scientific explanations. I am not suggesting (2) is likely false. I just try to emphasize that there are actually two quite different assertions (1) and (2) involved in the claim 'Gettier cases are not cases of genuine knowledge' and that only (1) is obvious.

Now, the epistemological intuition that I consider here is the intuition for judging (1), which is an instance of complex pattern recognition. As for (2), I agree that so far it is still unclear, and we need extensive empirical investigations on human cognitive activities in order to get a scientifically reasonable view regarding it. No mere intuition is sufficient to know (2). The intuition for judging (1) *is* a little trivial, but perhaps we anyway shouldn't expect that a mere intuition can achieve deep knowledge. This paper just claims that there *are* justifiable but practically irreplaceable intuitions for achieving knowledge, even though they are only for achieving some superficial knowledge. This epistemological intuition is anyway needed for recognizing the pattern K and for investigating the more challenging question (2), no matter if (2) turns out true or false. That is, it may be a little superficial, but it is *not* completely superfluous.

After these clarifications and qualifications, I hope it is clear that the epistemological intuition for judging Gettier cases *is* justifiable. More specifically, suppose that you present a new Gettier style case to an epistemologist and she judges that that new case does *not* instantiate the pattern K. Here she employs that epistemological intuition (as a complex pattern recognition ability). We notice that epistemologists are quite consistent

among themselves on such judgments. We have little idea about the neural mechanism realizing such a complex cognitive ability. However, given that the ability seems stable and consistent among a class of well-trained, intelligent people, and given our general scientific evidences supporting an overall naturalistic view about humans, we have reasons to believe that there *is* such a neural mechanism. It may not be an innate, special-purpose neural mechanism, like the facial recognition module. It is likely a combination of some other neural mechanisms formed after intensive trainings. Nevertheless, given the reliability of this intuition and given the scientific evidences that it is at least realizable by a neural mechanism, we *can* accept it as a justified method for achieving knowledge.

On the other side, epistemologists are already quite aware of the difficulty of analyzing the pattern K in explicit language. For instance, proposals for analyzing K into JTB + X for some X are so far unsuccessful. The situation appears to be similar to the situation in recognizing angry appearances. That is, one can get a *partial* description of K as JTB + X for some X, but a *complete* characterization of K itself seems inaccessible (no matter if K is identical with JTB + X for some X). Besides, it is not true that having a complete characterization of K is a precondition for our exercises of that epistemological intuition to be justified for judging that a case does instantiate the pattern K. On the contrary, we rely on that intuition to judge whether a proposed analysis of K into JTB + X is correct. Therefore, it seems we can agree that our exercises of this epistemological intuition are *practically* irreplaceable with any other methods, not to mention ordinary perceptual abilities and other hard-scientific methods.

Some other intuitions in philosophy appear to be similar to this. That is, given a philosophical claim allegedly made by using an intuition, we can sometimes identify a reliable assertion made by using our pattern recognition ability, like (1), and distinguish it from other less certain or even quite controversial assertions like (2). Then, if we carefully treat only the former as the real claim made by using the intuition (and agree that the latter actually needs extensive empirical investigations), we can say that that intuition is a justifiable method for delivering genuine knowledge, although after this qualification the knowledge it delivers becomes less significant. The naturalist should have no regret about this consequence, since the naturalist always takes it for granted that deep knowledge has to come by through extensive empirical investigations, not a mere intuition.

5 Naturalism and Irreplaceable Methods

Now consider the question whether admitting irreplaceable methods compromises the spirit of naturalism. We have seen that, under physicalism, the existence of irreplaceable cognitive methods is a physical (or actually, neural) fact about human brains. It is therefore completely compatible with physicalism as strong ontological naturalism.

Moreover, admitting irreplaceable cognitive methods does not degrade the special status of hard-scientific methods. We all agree that whenever hard-scientific methods *are* applicable, they are always more preferable. Scientific methods are more reliable and can lead to more exact descriptions, predictions and explanations. They are also more systematic and generalizable. That is, the same scientific methods can achieve knowledge in almost all areas, while our complex pattern recognition abilities are only

special-purpose tools (e.g., for recognizing human faces, angry appearances, etc.). Furthermore, scientific methods can reliably reach facts remote from us, both spatially (i.e., the micro and the astronomical) and temporally (i.e., the long past and future), and they can reach facts not accessible by our direct observations. In contrast, our pattern recognition abilities can reach only immediately observable appearances.

More importantly, there is a *physicalistic* explanation about why our complex pattern recognition abilities (and intuitions and mental simulations based on them) and scientific methods have such different features. For instance, consider the fact that mental simulations lack exactness and stability. To take a concrete example, consider again the mental simulation of a rock rolling down a hill (Williamson 2007, p.142). Suppose that you observe the initial movement of the rock, and then, in a mental simulation, you imagine a mental image to predict how the rock will continue to roll down at later moments, and you recognize that the image is sufficiently similar to your memories about real events of objects rolling down slanted planes. You can use this mental simulation to predict how the rock will really roll down the hill. Your neural mechanism realizing the mental simulation can do these because of the long history of adaptation to human environments.

However, the mental image you create for the simulation is quite fuzzy and unstable. For instance, it does not represent the *exact* spatial or temporal distance between two points in the image, and it may change slightly when you recall and reflect upon it later. Therefore, you cannot have an *exact* prediction based on this mental simulation. This lack of exactness and stability has many undesirable consequences. For instance, you cannot have exact tests about whether a mental simulation is accurate, and you cannot significantly improve your skills for creating more and more accurate mental simulations by practices.

This ability of mental simulation was directly endowed to us by evolution. The pressure of evolutionary adaptation demanded quick, easy, and immediate solutions to the problems we faced. The fuzzy, unstable image in a mental simulation was likely such a result of the demand for quick responses in pattern recognition tasks. It allowed the brain to ignore details that were not critical for human survival, recognize a pattern roughly but quickly, and then react to the environment promptly. This explains why our mental simulations lack exactness and stability.

We can compare this with the way our scientific methods solve the same problem. Scientific methods also create representations of real events in the world and use the representations for prediction. Such representations can also be considered images, that is, geometrical figures. For instance, a differential equation describing how an object moves is a rule for drawing a curve (as a solution to the equation) in a 4-dimentional Newtonian space-time reference frame to represent the movements of the object. However, here the image (namely, the curve and the reference frame) is build out of *exact, rigid* components (i.e., volume-less points, and rigid and strictly straight lines with no widths, and so on) in an *exact and stable* way. Our brains do not realize such an image directly in our visual cortex. Instead, we use the mathematical language to *describe* such image components and *describe* how an image is composed of such components. In other words, we ‘create’ the image by using our mathematical language, including our ability to follow rules in using the language. Note that ‘creating images’ is only a manner of speech here. What really happen in scientists’ brains are just some uses of the mathematical language, which we figuratively describe as

'creating (geometrical) images' from the subjective perspective. There is literally no image created.

This way of 'creating an image' is more complex and more indirect than, for instance, directly realizing a visual image as neural circuitries in the visual cortex. However, somehow it can 'produce' more exact, rigid, and stable images. This allows us to test an exact, stable image against what really happen in the real world by designing various scientific experiments. It also allows us to summarize the exact principles for drawing correct images to represent real events in the world, which are just laws of physics stated in the mathematical language, such as a differential equation describing the movements of an object.

Moreover, this way of creating representations (or rigid, exact, and stable images) is systematic and generalizable. We can consciously use those very exact basic image components and image constructing methods (namely, mathematical objects such as points, lines, and numbers, and mathematical model-construction methods) to build exact, stable, but very complex images to represent all kinds of things and events in the world. Exactness and rigidity of an image also allows it to represent things remote from us. In contrast, evolution can explain why our pattern recognition abilities and mental simulations lack such systematicity and generalizability. For instance, evolution is mostly ad hoc, addressing one adaptation requirement at a time. That is why our complex pattern recognition abilities are mostly special purpose tools and lack systematicity. Different neural mechanisms were developed for recognizing different patterns that were critical for our survival, to meet different adaptation demands. Occasionally, they *can* be reused for other purposes, but there is no *systematic* way for recombining and reusing various such abilities. Moreover, evolution can only select neural mechanisms that respond to immediately observable environmental features. That is why our pattern recognition abilities cannot reach remote or hidden facts.

Scientific methods were very late developments in human brains in the history of human evolution. Using the mathematical language to 'create' exact, stable but very complex images to represent things and events in the world is perhaps the greatest invention of human brains'. It is a brand new way to adapt to environments by human brains. It goes far beyond directly using our visual cortex to create visual images to represent environmental features and changes or directly using our other special-purpose neural modules (e.g., the facial recognition module) for recognizing more complex patterns in the environments.

On the other side, under physicalism, we can also explain why scientific methods are *not* very good at resolving *some* types of problems, for instance, human facial recognition and emotion recognition, where our innate complex pattern recognition abilities can perform much better. Because, the special-purpose neural mechanisms for performing such cognitive tasks are results of an extremely long history of evolution and have been deeply adapted to some very complex patterns in human environments. As a consequence, those special-purpose neural mechanisms can recognize those patterns very quickly and effortlessly. The patterns are very complex, and *consciously* creating 'mathematical images' *from scratch* to represent them will require a tremendous amount of explicit linguistic computations and inferences and is practically impossible for a brain.

Therefore, physicalism can explain *both* the existence of practically irreplaceable methods *and* their idiosyncratic features that make them different from hard-scientific

methods. It can explain *both* the advantages of hard-scientific methods in many situations *and* why in *some* situations we have to rely on our practically irreplaceable complex pattern recognition abilities, rather than hard-scientific methods. In this sense, admitting irreplaceable methods does not compromise the spirit of naturalism, even that of extreme naturalism. What the argument in this paper really shows is that a strong but coherent naturalistic worldview should consist of physicalism (as ontological naturalism) plus some *moderate*, rather than extreme, version of methodological naturalism.

Acknowledgments I am very grateful to the anonymous reviewer of *Philosophia*. The comments are very helpful. An earlier version of this paper was presented at *International Conference on Williamson, Logic and Philosophy, Peking University, October 16-17, 2015*. I wish to thank the participants for their comments.

References

Goldman, A. I. (2010). Philosophical naturalism and intuitional methodology. *Proceedings and Addresses of the American Philosophical Association, 84*, 115–150.

Haug, M. C. (Ed.) (2014). *Philosophical Methodology: The Armchair or the Laboratory?* Abingdon: Routledge.

Papineau, D. (2007). 'Naturalism', In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/naturalism/

Pust, J. (2012). 'Intuition', In E. N. Zalta (ed.) *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/intuition/

Rosenberg, A. (2014a). Why I am a naturalist. In M. C. Haug (Ed.), *Philosophical Methodology: The Armchair or the Laboratory?* (pp. 32–35). Abingdon: Routledge.

Rosenberg, A. (2014b). Can naturalism save the humanities? In M. C. Haug (Ed.), *Philosophical Methodology: The Armchair or the Laboratory?* (pp. 39–42). Abingdon: Routledge.

Stoljar, D. (2009). Physicalism, In E. N. Zalta (ed.), *Stanford Encyclopaedia of Philosophy*, http://plato.stanford.edu/entries/physicalism/

Symons, J. (2008). Intuition and philosophical methodology. *Axiomathes, 18*, 67–89.

Vaidya, A. J. (2010). Philosophical methodology: the current debate. *Philosophical Psychology, 23*, 391–417.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.

Williamson, T. (2014a). What is naturalism? In M. C. Haug (Ed.), *Philosophical Methodology: The Armchair or the Laboratory?* (pp. 29–31). Abingdon: Routledge.

Williamson, T. (2014b). The unclarity of naturalism. In M. C. Haug (Ed.), *Philosophical Methodology: The Armchair or the Laboratory?* (pp. 36–38). Abingdon: Routledge.