

**This item is the archived preprint of:**

Experimental philosophy and naturalism

**Reference:**

Nanay Bence.- Experimental philosophy and naturalism

Experimental Philosophy, Rationalism and Naturalism. Rethinking Philosophical Method / Fischer, Eugen [edit.]; et al. - ISBN 978-1-138-88728-2 - London, Routledge, 2015, p. 222-239

## EXPERIMENTAL PHILOSOPHY AND NATURALISM

Bence Nanay

The aim of this paper is to argue that there has been some mismatch between the naturalist rhetoric of experimental philosophy and its actual practice: experimental philosophy is not necessarily, and not even paradigmatically, a naturalistic enterprise. To substantiate this claim, a case study is given for what genuinely naturalist experimental philosophy would look like.

### I. Experimental philosophy and ‘intuitions’

Many philosophers are attracted to experimental philosophy because of its naturalist appeal: on the face of it (more about this in the next section), naturalism entails that philosophy should be continuous with the sciences and if philosophers do experiments, experimental philosophy seems to be an important move in the naturalist direction. The aim of this paper is to show that the relation between experimental philosophy and naturalism is much more complicated and the naturalist appeal of experimental philosophy is in many, even most, cases, misleading

After some clarifications on what I mean by experimental philosophy and naturalism (Section I and Section II, respectively), I argue that most experimental philosophy studies are not naturalistic at all (Section III) and I explore the ways in which experimental philosophy can be naturalized (Section IV).

Following Prinz (2008), I make a distinction between experimental and empirical philosophy. Empirical philosophy uses empirical findings in order to give an account of some philosophically interesting aspect of the world. Experimental philosophy differs from empirical philosophy in two ways: first, it does not merely use others’ experimental findings, but also conducts experiments, and, second, these experiments are not about some philosophically interesting aspect of the world, but about some philosophically interesting aspect of the way we think about the world. They are not experiments about causation, say, but about our intuitions about, or our conception of, causation (Hitchcock & Knobe 2009, Knobe & Fraser 2008).

Results about our intuitions about a phenomenon do not automatically yield results about the phenomenon itself, as our intuitions can be, and very often it in fact are, misleading. I used, and will continue to use, the phrase ‘intuition’ but it needs to be acknowledged that the concept of intuition is by no means an unproblematic one (Cappelen 2012, Deutsch 2010, cp. Introduction to this volume) and I need to say something about the troubled relation between the concept of intuition and the experimental philosophy research program.

It seems that there are two very different general approaches within the tradition of experimental philosophy when it comes to the role intuitions play that are sometimes labeled the ‘positive view’ and the ‘negative view’. The first of these approaches takes experimental findings to ‘democratize intuitions’. The only way experimental philosophy differs from old-fashioned conceptual analysis is that it takes into consideration not only the intuition of one philosophically trained person, but that of hundreds of subjects. We can call this approach the non-radical version of experimental philosophy as it emphasizes the continuity with the old tradition of conceptual analysis.

The second, more radical approach (that is sometimes called the ‘negative view’) uses experimental philosophy to undermine the role of intuitions in philosophy altogether. This research program aims to point out important variations in our intuitions, which jeopardizes the whole idea of relying on intuitions (Weinberg et al. 2001, Swain et al. 2008, for a somewhat more moderate version, see Stich forthcoming and Weinberg, this volume). Intuitions change as a result of framing effects, order effects, environmental effects and demographic differences (Schnall et al. 2008, Petrinovich & O’Neill 1996, Tversky & Kahneman 1981, Haidt & Baron 1996, Machery et al. 2009, Machery et al. 2004, Swain et al. 2008, Valdesolo & DeSteno 2007, see also Sinnott-Armstrong 2006 and Stich forthcoming for summaries). But if intuitions vary this widely and if they depend on seemingly irrelevant parameters, then we have good reason not to trust them.<sup>1</sup>

These radical conclusions have been questioned in various different ways, both from naturalist (cp. Kornblith, this volume) and from autonomist perspectives: Maybe the variation of intuitions is not such a big problem as long as these variations are only on the surface and there is some deep common denominator between them (Kauppinen 2007, see also Sosa 2006). More importantly, it could be argued that those who push or these radical conclusions in fact use reliable intuitions for establishing the variations of our intuitions, thereby defeating their own claim (Liao 2008).

Without analyzing these arguments in detail, it needs to be noted that we have independent (and not particularly novel) reasons for mistrusting the reliability of our intuitions. In fact, it would be extremely surprising if all our intuitions were reliable in all contexts. Our minds evolved to get most things approximately right most of the time. They did not evolve to get all things completely right all the time. Take folk chemistry. Most of us would say that if we add some transparent liquid to another transparent liquid, we will get a transparent liquid. This is true in the overwhelming majority of cases. But not always. If we add some water to a glass of pastis, we get an opaque liquid (the same goes for ouzo, sambucca, raki, etc). And folk chemistry is not an isolated example. We have misleading intuitions about what we see. As the inattentive blindness literature shows, we are aware of a very small proportion of the features

---

<sup>1</sup>Other approaches derive such negative assessments from psychological explanations of our intuitions, rather than from sensitivity to irrelevant parameters. Cp. Fischer, this volume.

of the visual scene in front of us (Simmons-Chabris 1999, Mach-Rock 1998, Nanay 2010a). Nonetheless, most of us would say that we are fully aware of the entire scene we are looking at. As folk chemistry is a bad guide to the truths about the chemical composition of the world, folk psychology is also a bad guide to the truths about how the mind works.

In general, we should not infer ‘P’ from ‘I have the intuition that P’ and we should not infer ‘P’ from ‘hundreds of people have the intuition that P’ either. I do not take this to be a particularly controversial claim and, importantly, I do not think that either classical conceptual analysis or the non-radical version of experimental philosophy routinely rely on inferences of this kind. The question I would like to explore is about the nature of the relation between these two claims, that is, ‘hundreds of people have the intuition that P’ and ‘P’, in the framework of experimental philosophy. The experiments will deliver claims of the former kind. The question is: How can and should these claims be used to arrive at claims of the latter kind? And this brings us to the question of naturalism.

## **II. What is naturalism?**

The most important naturalist slogan since Quine has been that philosophy should be continuous with science. As he says, “I admit to naturalism and even glory in it. This means banishing the dream of a first philosophy and pursuing philosophy rather as a part of one’s system of the world, continuous with the rest of science” (Quine 1984, pp. 430-431). Unfortunately, it is not entirely clear what this is supposed to mean.

One thing it may mean is that philosophy should not postulate the existence of supernatural entities. This is not a sense of naturalism that I am interested in here. This concept of metaphysical naturalism is often contrasted with a methodological one: philosophy should use similar methods as science uses. This is a more relevant notion, but we need to proceed with care to see what this entails. Those who interpret the research program of experimental philosophy as naturalist presumably have a version of this interpretation in mind. Both scientists and experimental philosophers do experiments – they have the same methodology then, which makes the endeavor of experimental philosophy naturalistic. But it is important to note that this way of thinking about naturalism will not do. We can at least in principle use empirical results or even conduct experiments in order to try to establish that, say, the mind cannot be explained in a naturalized manner. Hence, similarity of methodology between the sciences and philosophy does not guarantee naturalism. For a coherent account of what is meant by naturalism, we need to look further.

I will examine the question of naturalism in the context of philosophy of action and then generalize the results to philosophy in general (see also Nanay forthcoming). I chose philosophy of action because there has been a lot of discussion in the last three decades on what shape or form naturalized philosophy of action may take, and also because the experimental philosophy of action constitutes a particularly well-developed subfield of experimental philosophy.

The most important proponent of the naturalization of action theory was Myles Brand. He argued that philosophy of action should enter into its ‘third stage’ (the first one was the 1950s and 60s, the second the 70s), the main mark of which would be its continuity with the empirical sciences. This third stage, the stage of naturalized action theory did not happen. The last thirty years of philosophy of action could be described as doing the exact opposite of what Brand suggested. Contemporary philosophy of action is not a naturalist enterprise: it consistently ignores empirical findings about actions and its mental antecedents: it has no patience for the cognitive neuroscience of action, for example. But the debate following the publication of Brand’s book will help us to clarify what would make for a viable and yet nontrivial conception of naturalism about action theory.

Naturalism in the context of philosophy of action can be, and has been, formulated in a Quinean manner. Brand himself, for example, talks about “the integration of the philosophical with the scientific” in the domain of action theory (Brand 1984, p. x).

Just what this ‘continuity’ or ‘integration’ is supposed to mean, however, remains unclear. More specifically, what happens if what science tells us is in conflict with what folk psychology tells us? Brand clearly hands the decisive vote to folk psychology. As he says, “Scientific psychology is not free to develop any arbitrary conceptual scheme; it is constrained by the conceptual base of folk psychology” (Brand 1984, p. 239). But that has little to do with naturalism, as Slezak 1987 and 1989 points out (see esp. the detailed point by point analysis of how Brand’s theory fails on its own terms in Slezak 1989, pp. 140-141, pp. 161-163). If the only role science is supposed to play in action theory is that it fills in the details of the already given, not changeable conceptual framework of folk psychology, then science is not playing a very interesting role at all – the conceptual framework of action theory would still be provided by folk psychology. Brand’s theory, is not naturalistic.

What would then constitute a naturalized action theory? We can use Brand’s original formulation as a starting point: naturalized action theory urges the integration of the philosophical with the scientific, but a very specific kind of integration: one where the philosophical does not automatically trump the scientific. If it turns out that some of our key concepts in philosophy of action (like those of ‘action’ or ‘intention’) fail to pick out any natural kinds, we have to replace them with concepts that do pick out natural kinds. And science can tell us what this new concept is supposed to be.

I talked about the importance of empirical findings in naturalized action theory: empirical findings constrain the philosophical theories of action we can plausibly hold. But the interaction between philosophy and the empirical sciences is bi-directional. The philosophical hypotheses and theories, as a result of being empirically informed, should be specific enough to be falsified or verified by further empirical studies. Psychologists and neuroscientists often accuse philosophers in general, and philosophers of mind in particular, of providing theories that are too general and abstract: that are of no use for the empirical sciences. Naturalized action theory would correct this mistake and give exact, testable hypotheses that psychologists as well as cognitive neuroscientists of action can engage with. Naturalized action theory,

besides using empirical studies, could also be used for future empirical research. This is the only sense in which the ‘integration of the philosophical with the scientific’ that Brand talked about does not become a mere slogan.

And what goes for philosophy of action also goes for philosophy in general. What it means for a philosophical research program to be naturalist is that it integrates the philosophical and the scientific in such a way that the philosophical does not automatically trump the scientific. A naturalistic approach is always open to re-evaluating our concepts and intuitions in the light of stronger scientific evidence. And naturalistic accounts are not only informed and constrained by scientific findings, but they are also, at least in principle, in the position to be relevant to further scientific research. Now finally we are in the position to ask how experimental philosophy fares in this respect.

### III. Is experimental philosophy naturalistic?

On the face of it, experimental philosophy is not doing so well when it comes to this criterion of naturalism. Experimental philosophy, as we have seen in Section I, conducts experiments about our folk intuitions. Thus, on the basis of these experiments, we can learn about our folk intuitions – and not directly about what these folk intuitions are about.

This is the contrast with empirical philosophy, which uses scientific findings about a phenomenon when formulating a philosophical account of this phenomenon. Call the phenomenon in question X. In this case, the real question is about the relation between the following:

- i. Experimental findings concerning intuitions about X
- ii. Philosophical account of X
- iii. Scientific findings about X

Empirical philosophy takes the relation between (ii) and (iii) seriously. It gives a philosophical account of X that takes the scientific findings about X to inform and maybe constrain these philosophical theories. Note that this does not necessarily imply naturalism. This interaction between (ii) and (iii) can be a one-sided one: The philosophical theory of X can be sensitive to the scientific findings concerning X, but if this philosophical theory then is completely irrelevant to any further scientific study of X, then the project is not a naturalistic one. Empirical philosophy may or may not be naturalistic.

How about experimental philosophy? If empirical philosophy was about the relation between (ii) and (iii), then experimental philosophy is about the relation between (i) and (ii). It is important to emphasize that this relation can take many forms. We have seen that taking (i) to be the evidence for (ii) is problematic in its simplest form. Experimental findings concerning our intuitions about X do not necessarily deliver the correct results about X – they

would only do so if intuitions were completely reliable and as we have seen, we have plenty of reasons to mistrust our intuitions.

An interesting way of approaching this relation between (i) and (ii) is Joshua Knobe's. He uses experimental results concerning our intuitions about X to draw some philosophical conclusions not about X, but about a different phenomenon, Y. Probably the most famous example is this: His starting point is the experimental finding that the attribution of intentionality is influenced by moral considerations (Knobe 2003, 2007). This is an experimental finding about our folk intuitions: about which actions *we take to be* intentional (not about which actions *are* intentional). But he does not stop at this result but asks: what about the functioning of our mind makes it the case that the attribution of intentionality is influenced by moral considerations (see esp. Knobe 2010). And he examines two possible explanations. The first one is that the competences that underlie our mental capacities (to attribute intentionality or to spot causal relevance) are influenced by moral considerations. The second is that these competences are themselves non-moral, but there is some additional factor that makes it the case that our attribution of intentionality is influenced by moral considerations (see also Machery 2008, Mallon 2008, Nanay 2010b, Nanay 2010c, Nichols-Ulatowski 2007, Hindriks 2008). When deciding between these two (and settling for the former view), he uses further experiments, but these experiments are also about our folk intuitions. Crucially, at no point of this analysis are these explanations of folk intuitions subject to revision in the light of stronger scientific findings, say, from the cognitive neuroscience of intentional action. In this sense, Knobe's approach is not naturalistic.

But some other approaches to experimental philosophy studies can be naturalistic. The radical version (or, the 'negative program') of experimental philosophy I mentioned in Section I uses experiments about our folk intuitions in order to show that we are not justified to accept some of our folk intuitions. So they connect (i) and (ii) in a very different way from other experimental philosophers. They use experimental findings about our intuitions concerning X to say something about our philosophical accounts *of the intuitions* concerning X. Note that this approach is not about the relation between experimental findings about our intuitions concerning X and the philosophical accounts of *X itself*, but about the relation between experimental findings about our intuitions concerning X and the philosophical accounts *of our intuitions concerning X*. In this sense, this approach uses experimental findings to inform philosophical accounts, but both the experimental findings and the philosophical accounts are about our intuitions concerning X and not about X, the phenomenon itself. The negative program of experimental philosophy, or at least some versions thereof (e.g. Fischer, this volume), then can be considered to be naturalistic, but with an important *caveat*: it can give rise to a naturalistic philosophical account of our intuitions concerning X, but that is, of course, different from a naturalistic philosophical account of X itself. To use one of the main areas of research of the proponents of the negative view, if we take X to be 'knowledge', what the negative view achieves is a naturalistic account of our intuitions concerning knowledge, not a

naturalistic account of knowledge itself. This is naturalism of a very different kind from the naturalism empirical philosophy can achieve.

Should we then conclude that while experimental philosophy can be naturalistic, it can only provide a naturalistic philosophical account of our intuitions concerning X and not a naturalistic philosophical account of X itself? No, we shouldn't. Further, those who mistrust the radical branch of experimental philosophy (maybe for some of the reasons I mentioned in Section I), do not have to give up on the naturalist project altogether. It is possible to use experimental philosophy in a naturalist, yet, non-radical way. In order to see how this may work, I give a case study of a naturalist but non-radical experimental philosophical study in the next section.

#### **IV. A case study: experimental philosophy and the concept of action**

A lot of work in experimental philosophy has focused on the difference between intentional and unintentional actions. I want to examine an even more fundamental question in action theory: what makes actions *actions*. One of the most important tasks of action theory is to tell us the difference between actions and mere bodily movements: to tell us what makes action actions (and not mere bodily movements).

Some bodily movements are actions. If I decide to eat some yoghurt and get up from my computer to do so, I perform an action. If my leg moves because the doctor is examining my knee-jerk reflex with her hammer, I do not perform an action – we have to talk about mere bodily movement. One of the most important tasks of action theory is to tell us the difference between actions and mere bodily movements: to tell us what makes action actions (and not mere bodily movements).

Most accounts of what makes actions *actions* point to some kind of mental state that triggers (or maybe accompanies) our bodily movement. If the bodily movement is triggered (or maybe accompanied) by this specific kind of mental state, it is an action. If it is not, it is a mere bodily movement.

The big question is of course what mental states are the ones that trigger (or accompany) actions. And there is no consensus about what these mental antecedents of actions are supposed to be. Myles Brand called mental states of this kind 'immediate intentions' (Brand 1984), Kent Bach 'executive representations' (Bach 1978), John Searle 'intentions-in-action' (Searle 1983), Ruth Millikan 'goal state representation' (Millikan 2004, Chapter 16), Marc Jeannerod 'representation of goals for actions' and 'visuomotor representations' (Jeannerod 1994, section 5, Jeannerod 1997, Jacob-Jeannerod 2003: 202-204).<sup>2</sup> I myself called them 'action-oriented perceptual states' or 'pragmatic representations' elsewhere (Nanay 2011, 2012, 2013, forthcoming). Here, I will use the term 'the immediate mental antecedents of actions' as

---

<sup>2</sup> This list is supposed to be representative, not complete. Another important concept that may also be listed here in John Perry's concept of 'belief-how' (Israel et al 1993, Perry, 2001).

a place-holder for whatever mental state makes actions *actions*. I need to emphasize that although I will talk about the immediate mental antecedents of actions, what I say about them can be generalized to any kind of mental state that is supposed to make actions *actions*, regardless of whether they cause or accompany the bodily movement. The term ‘immediate mental antecedent of action’ is only supposed to be a convenient label.

An important clarification: the immediate antecedents of action are the immediate antecedents of all of our actions. Not just of intentional actions; all actions. And not just of autonomous actions (see Velleman 2000, Hornsby 2004), all actions. Thus, understanding the nature of the immediate mental antecedents of actions is a logically prior task for philosophers of action to all other questions in action theory. In order to even set out to answer questions like ‘what makes actions intentional?’ or ‘what makes actions autonomous?’, one needs to have an answer to the question ‘what makes actions *actions*?’.

In the case of these two examples mentioned above (getting up to eat yoghurt and having a knee-jerk reflex), it is clear enough whether we should count them as actions or mere bodily movements. If an account of what makes actions *actions* classifies knee-jerk reflexes as genuine actions, we have reason to be suspicious. But we can use the method of experimental philosophy to show that there are some examples that do not fall so clearly into one of these two categories. These experiments demonstrate that action-attribution has some important intermediary cases and, as a result, any serious account of what makes actions *actions* needs to give an explanation of why we have conflicting intuitions in these intermediary cases – I call these intermediary cases, examples that are neither fully-fledged actions nor mere bodily movements, ‘semi-actions’.

Consider the following six examples:

- (a) I decide to eat some yoghurt and get up from my computer to do so.
- (b) While typing at my computer, I suddenly and without planning to do so, jump up and start pacing around the room (the example is from Searle 1983).
- (c) I stand behind a strong piece of Plexiglas knowing that there is Plexiglas in front of me, when someone on the other side of the glass throws a beach ball at me and I reach out in an attempt to catch the ball (the example is from Nanay 2012).
- (d) “I put my face close to the thick glass-plate in front of a puff-adder in the Zoological Gardens, with the firm determination of not starting back if the snake struck at me; but, as soon as the blow was struck, my resolution went for nothing, and I jumped a yard or two backwards with astonishing rapidity.” (Darwin 1899, p. 18.)
- (e) Anarchic hand syndrome patients exhibit clearly goal-directed, well-executed and complex movements of an upper limb that are nonetheless unintended and sometimes conflict with the agent’s desires or even the action they intentionally perform with their other hand (Della Sala et al. 1991, 1994, Giovannetti et al. 2005, Kritikos et al. 2005, see also Pacherie 2007 for a philosophical summary).

- (f) The doctor is examining my knee-jerk reflex with a hammer and my leg moves as a result.

We have already seen that (a) is a clear example of an action. In fact, it is an *intentional* action (thankfully, I can ignore in this context what that means). Case (b) is also widely held to be an action, although opinions differ about whether it is an intentional action. Case (f) is widely held to be a mere bodily movement – I do not perform any action.

The problem is with the remaining three examples, where it seems that we have no principled way of telling whether they are actions or not. None of them seems to be an intentional action, but this leaves the question about whether they are (non-intentional) actions or mere bodily movements unanswered.

Cases like (c), (d) and (e) are not so easy to categorize. My suggestion is that it is problematic to consider them to be either fully fledged actions or mere bodily movements. They fall somewhere in between. Luckily, our intuitions about these cases can be examined experimentally. And it turns out that while people have no problems categorizing cases like (a) and (b) as actions and (f) as non-action, they are torn when they are asked about cases like (c), (d) and (e). I demonstrated this with classical experimental philosophy methods: subjects were presented with the description of a bodily movement and then asked to agree or disagree (on a scale from 1 (strongly disagree) to 7 (strongly agree)) that in this scenario *‘I am performing an action’*.

*First experiment:*

Three scenarios (which are all variations of (c) above) were tested:

- (g) Someone throws a beach ball at me and as I want to catch it, I reach out in an attempt to do so.
- (h) I made a bet with my friend to stay motionless for 2 minutes. But then she throws a beach ball at me and although I do not want to move, when it is flying towards me, I do reach out to catch it.
- (i) I have an odd muscle contraction and my hands move towards the ball just as it is flying towards me.

Assuming that (g) is a genuine action (mean: 6.49) and (i) is a mere bodily movement (mean: 4.46), (h) (mean: 5.80) was significantly different both from (g)<sup>3</sup> and from (i)<sup>4</sup>.

The mean in the case of (i) is much higher than one would expect (given that value 1 is ‘strongly disagree that ‘I’m performing an action’ and value 7 is ‘strongly agree that ‘I’m

---

<sup>3</sup> Results were subjected to an independent samples T-Test:  $t(123) = 3.15$ ,  $p = 0.002$  (two-tailed),  $SD$  ((g)) 0.68,  $SD$  ((h)) 1.50, Cohen’s d: 0.59

<sup>4</sup>  $t(126) = 4.46$ ,  $p = 0.000$  (two-tailed),  $SD$  ((i)) 1.89,  $SD$  ((h)) 1.50, Cohen’s d: 0.78

performing an action". But it is important to note that all non-actions scored relatively high on this scale – the kneejerk reflex scored the mean of 3.81, for example. And the fact that the mean of semi-actions is significantly different from the mean of genuine actions as well as this relatively high mean of non-actions shows the robustness of the findings.

*Second experiment:*

Three scenarios (which are all variations of (d) above) were tested:

- (j) I am looking at a dangerous snake from close range, and as the snake strikes, I don't want to be bitten so I jump back.
- (k) I put my face close to the thick glass-plate in front of a snake, with the firm determination of not starting back if the snake struck at me; but, as soon as the blow is struck, I jump a yard or two backwards with astonishing rapidity.
- (l) I put my face close to the thick glass-plate in front of a puff-adder and as the snake strikes, the zoo personnel throws me back a yard or two.

Assuming that (j) is a genuine action (mean: 5.80) and (l) is a mere bodily movement (mean: 3.48), (k) (mean: 5.13) was significantly different both from (j)<sup>5</sup> and from (l)<sup>6,7</sup>.

*Third experiment:*

Three scenarios (which are all variations of (e) above) were tested:

- (m) Anarchic hand syndrome (AHS) patients exhibit clearly goal-directed, well-executed and complex movements of an upper limb that are nonetheless unintended and

---

<sup>5</sup>  $t(123) = 2.24, p = 0.024$  (two-tailed),  $SD$  ((j)) 1.52,  $SD$  ((k)) 1.80, Cohen's d: 0.40

<sup>6</sup>  $t(129) = 4.79, p = 0.000$  (two-tailed),  $SD$  ((l)) 2.07,  $SD$  ((k)) 1.80, Cohen's d: 0.85

<sup>7</sup> One might object: (h) and (j) have the same mean value. But one, (j), is characterized here as a genuine action, whereas the other is taken to be a semi-action. How is this possible? The answer is that the methodology used here is not one where all bodily movements are placed on a scale from (1) to (7) depending on how much we consider them to be actions and then look for three clusters corresponding, respectively, to genuine actions, to mere bodily movements and to semi-actions somewhere between the two. Rather, the methodology is to compare like with like. So, according to an independent samples T-Test, (h) is significantly different from both (g) and (i), where (g), (h) and (i) are all part of the same scenario of catching a ball (and the bodily movement is also the same). And assuming that (g) is a genuine action and (i) is a mere bodily movement, this means that (h) should be categorized as neither – as a semi-action. Different scenarios with different bodily movements (for example Darwin's snake scenario) may (and do) have much lower means for all three values (for what counts as genuine action, for what counts as mere bodily movement and for what counts as semi-action).

sometimes conflict with the agent's desires. Suppose that I am not an AHS patient and start strangling myself with my left hand.

- (n) Anarchic hand syndrome (AHS) patients exhibit clearly goal-directed, well-executed and complex movements of an upper limb that are nonetheless unintended and sometimes conflict with the agent's desires. Suppose I am an AHS patient and start strangling myself with my left hand.
- (o) Anarchic hand syndrome (AHS) patients exhibit clearly goal-directed, well-executed and complex movements of an upper limb that are nonetheless unintended and sometimes conflict with the agent's desires. Suppose that I am not an AHS patient and you grab my hand and start strangling me with it.

Assuming that (m) is a genuine action (mean: 5.61) and (o) is a mere bodily movement (mean: 2.99), (n) (mean: 3.99) was significantly different both from (m)<sup>8</sup> and from (o)<sup>9</sup>.

These results show that our folk intuitions discern three categories: actions, semi-actions and mere bodily movements. But most philosophical accounts of what makes actions *actions* rely on the assumption that there are only two such categories: actions and mere bodily movements. In other words, there is a conflict between our folk intuitions, as revealed by these experiments, and our philosophical accounts of action.

To use the tripartite distinction I made in Section III, we have a conflict between (i) and (ii):

- i. Experimental findings concerning intuitions about X
- ii. Philosophical account of X
- iii. Scientific findings about X

How should we resolve this conflict? One way we could resolve it would be to insist, maybe with some backing from the negative program of experimental philosophy, that our intuitions in this case are unreliable: we may have the folk intuitions that semi-actions are different both from *bona fide* actions and from mere bodily movements, but in fact, they are not. In this case, we can continue working with the assumption that there is a mental state type that makes actions *actions* – the problem of semi-actions is avoided. This solution would resolve the conflict between (i) and (ii) by throwing out (i).

Another way of resolving the problem would be to reject those philosophical theories of what makes actions *actions* that rely on the assumption that there is one and only one mental state type that makes actions *actions* as accounts that blatantly wouldn't allow for the existence

---

<sup>8</sup>  $t(238) = 7.16, p = 0.000$  (two-tailed),  $SD((m)) 1.66, SD((n)) 1.83$ , Cohen's d: 0.92

<sup>9</sup>  $t(243) = 4.00, p = 0.000$  (two-tailed),  $SD((o)) 2.04, SD((n)) 1.83$ , Cohen's d: 0.51

of semi-actions. This way of resolving the conflict between (i) and (ii) would do so by favoring (i) over (ii).

It seems that either (i) or (ii) has to go. What makes actions *actions* is that they are triggered (or accompanied) by a mental state of a certain kind. If they are triggered (or accompanied) by such a mental state, they count as actions. If they are not, they are mere bodily movements. There doesn't seem to be any logical space for a middle way in between.

But there is a third possible approach, namely bringing in (iii). That is, the conflict between the experimental findings about our folk intuitions concerning action and our philosophical accounts of action could be resolved in such a way that we do not need to reject either (i) or (ii) as long as we take the scientific findings about action seriously. The suggestion then is that we can develop an account about the immediate mental antecedent of actions that can explain why we hesitate to characterize (c), (d) and (e) either as actions or as mere bodily movements. In order to do so, however, we need to rely on an important distinction from the cognitive neuroscience of action.

Some contemporary cognitive scientists make a distinction between two constituents of the immediate mental antecedents of action: one that represents the world (or the action's goal) in such a way as to be able to guide the action and one that moves us to act (see, for example, Haggard 2005a, esp. p. 293, see also Haggard 2005b, Jeannerod 1997 and also the literature on utilization behavior, see Lhermitte 1983, Shallice et al 1989, Pacherie 2007, Frith et al. 2000, Archibald et al. 2001, Shallice and Burgess 1991). This distinction is surprisingly similar to one that was made by Myles Brand in the early 1980s and then got largely forgotten (Brand 1984, p. 45). Brand calls these two constituents of the immediate mental antecedents of action (that he calls 'immediate intentions') the 'cognitive' and the 'conative' components. I will use the same terminology in what follows. The cognitive component of the immediate mental antecedent of actions is guiding and monitoring of ongoing activity, and it may or may not be conscious. According to Brand, this is necessary but not sufficient for acting. What is also needed is that we are 'moved to act'. This is the conative component of the immediate mental antecedent of action.<sup>10</sup> The general insight from cognitive science then is that the immediate mental antecedent of action has two distinct components: one that represents the world, or the immediate goal of the action, in a certain way, and the other one that moves us to act.

As long as we acknowledge that the immediate mental antecedents of action consist of two distinct components, we can give a coherent account of the intermediary cases of action attribution. The upshot is simple: typically, in the case of performing actions both components are present and in the case of mere bodily movements neither of them are. In the intermediary

---

<sup>10</sup> Kent Bach makes a similar (but not identical) distinction between 'receptive representation' and 'effective representation' that together make up the 'executive representation' that is the immediate mental antecedent of action (Bach 1978, see esp. p. 366).

cases where we have conflicting intuitions about whether they qualify as actions, only one of the two components (the cognitive one) is present. But let us proceed more slowly.

In (a) and (b), the immediate intention has both a cognitive and a conative component. The conative component in (b) is unplanned but nonetheless it is still there. And the cognitive component is also there: in (a) I need to represent the contents of the fridge (as well as the whereabouts thereof) and in (b) I need to have some kind of representation of where in my office one can pace up and down and when to turn around in order not to bump into the bookshelf. In the case of (f), in contrast, we have neither of these two components of the immediate mental antecedent of action: neither the cognitive, nor the conative one: we don't need to represent the world in any way in order for our knee-jerk reflex to work and what moves us to act is not the conative component of the immediate mental antecedent of action but the doctor's hammer.

In (c), (d) and (e), the conative component is missing, while the cognitive component is present. We are 'moved to act' by some external stimulus. This is what makes (c), (d) and (e) somewhat similar to reflex movements, where there is no conative component either. But, importantly, the cognitive component is there: the cognitive component of our immediate intention is guiding and monitoring our ongoing activity: that is why we reach out for the ball in (c) in a certain way that reflects the size of the ball and the direction it is coming from. And that is why Darwin jumps in a certain direction (and not others) in (d). Finally, this is the reason why anarchic hand patients can perform extremely complex goal-directed movements with their anarchic hand and do so successfully in (e).<sup>11</sup>

---

<sup>11</sup> This is an important contrast between what I call 'semi-actions' and what Brian O'Shaughnessy called 'subintentional acts' (O'Shaughnessy 1980, pp. 58-73). Tapping one's foot when listening to music is a subintentional act, but it is not a semi-action – you do not need to represent the world in any way in order to be able to tap your foot, whereas, as we have just seen, you do need to represent the world in a certain way in order to perform semi-actions. O'Shaughnessy's 'subintentional acts' could be considered to be 'mere bodily movements' (albeit, with the important feature that we can stop them if we want to), whereas this is not true of semi-actions. One way of understanding the difference between what I call semi-actions and what O'Shaughnessy calls 'subintentional acts' is that while in the case of semi-actions, the cognitive component of the immediate mental antecedent of action is present, the conative is missing, whereas subintentional acts could be interpreted as acts where the cognitive component is missing, while the conative is present – what moves me to tap my foot when listening to music is not something external – it comes from within. Some other potential examples (that O'Shaughnessy may or may not consider to be bona fide subintentional acts): blinking (without planning to do so and without any external influence that would trigger the blinking reflex) and swallowing (without planning to do so and without any external influence that would trigger swallowing the reflex). In these cases, we do not need to represent anything in order for an action to be performed. So the cognitive component is missing. But there is a conative component: whatever

## V. Conclusion: Naturalist versus non-naturalist experimental philosophy

In this case study, we found a conflict between the experimental findings concerning our intuitions about actions and our philosophical theories of action. This conflict was resolved with the help of some scientific findings about the two different constituents of the immediate mental antecedents of action. In other words, (iii) helped us to resolve the conflict between (i) and (ii). This is a genuinely naturalist explanatory scheme for an experimental philosophy project. But note that it is not of the kind the negative view of experimental philosophy advocates. This is naturalist but non-radical experimental philosophy.

It is not entirely straightforward whether and how this case study could be generalized. First, in the case of philosophy of action, there is a corresponding scientific discipline that is about the subject matter of the philosophical discipline: the cognitive neuroscience of action. Both philosophy of action and the cognitive neuroscience of action are about the same subject: action. But not all philosophical subfields are like that. In the case of epistemology, for example, it is not particularly straightforward to rely on scientific results concerning knowledge – as opposed to scientific results concerning our intuitions about knowledge, which belongs to (i) and not to (iii). And this is even more problematic when it comes to metaphysics. In the case of philosophy of action, we have both (i), (ii) and (iii) at our disposal: experimental findings about our intuitions, philosophical accounts and scientific accounts. But in some other subfields of philosophy, (iii) is missing: there are no direct scientific findings about the matters at issue in the three-dimensionalism versus four-dimensionalism debate or in the tropes versus universals debate. In these cases it may simply not be possible to run the version of naturalist experimental philosophy the present case study does. Note, however, that in these cases, it is also problematic to talk about empirical philosophy. As the explanatory scheme of naturalist experimental philosophy is supposed to unify experimental philosophy and empirical philosophy, this is exactly what we should expect.

Second, one may wonder what role (i) plays in this explanatory scheme. Is it more than heuristics? I am not sure. In the case study I presented, (i) gave us reasons to mistrust our philosophical theories of action and this could lead to bringing in (iii). One way of putting this is that (i) gave us a new explanandum: semi-actions. Semi-actions don't show up either in our naïve intuitions or in our philosophical accounts.<sup>12</sup> They do show up in (i): in the experimental

---

moves me to act is not something external (as in the blinking/swallowing reflex case), but something that comes from within. Do these actions count as *bona fide* actions? Do they count as semi-actions? I am not sure. What is more important than the labels is that the distinction between the cognitive and the conative components of the mental antecedents of action allows us to differentiate these importantly and interestingly different cases.

<sup>12</sup> Leaving aside O'Shaughnessy's related but, as we have seen in the previous footnote, very different, concept of subintentional acts.

findings about our intuitions concerning actions. It is possible that if we knew more about the cognitive neuroscience of action, a genuine empirical philosophy approach could have resulted in the acknowledgement of semi-actions and an explanation of it in terms of the two different constituents of the immediate mental antecedents of action. In this sense, the naturalist experimental philosophy approach I am advocating is closer to empirical philosophy than to experimental philosophy *tout court* – (iii) is the *sine qua non* of the naturalist approach, after all. But, and this is where taking (i), (ii) and (iii) into consideration pays off, (i) can, and, as the case study shows, it often does, help us to make the interaction between (ii) and (iii) more focused and it can even give us *explananda* that (ii) and (iii) need to work together to explain. Further, as experiments about our intuitions are much, much cheaper than fMRI experiments, we can use (i) to get a sense of what kind of (iii) studies we should conduct that could then be used fruitfully by (ii). Although the interaction between (ii) and (iii) is the defining feature of any naturalistic approach, triangulating between (i), (ii) and (iii) can make this naturalistic approach much simpler and more feasible.

Finally, it needs to be emphasized that while in the case study I presented, the conflict between (i) and (ii) was resolved by (iii) in a way that favored (i), this is by no means necessarily true. In the case study, the findings from the cognitive neuroscience of action were used to explain the importance of semi-actions – something we postulated on the basis of (i) and in spite of the theoretical commitments of (ii). Here, the experimental findings about our intuitions concerning action were winning out against the philosophical accounts. But this is not always, and maybe not even normally, the case. Often, as our intuitions are misleading, the experimental results about our intuitions will also be misleading and the scientific results (iii) can be used to explain why (ii) get things right and why (i) get things wrong. The triangulation between (i), (ii) and (iii) can go in a variety of ways. But as long as we have a bidirectional interaction between (ii) and (iii), and as long as (i) is also involved in some ways, we have a naturalistic version of experimental philosophy.

A final question: why should we prefer naturalist over non-naturalist versions of experimental philosophy? The short answer is that we should not. Both kinds of projects can be valuable and neither should be dismissed. But many, maybe even most, experimental philosophers use naturalistic rhetoric and present experimental philosophy as an important (maybe even the last) step towards establishing the continuity between philosophy and the sciences. This, if we take the rhetoric at its face value, is, in many cases, false advertisement. Those who are drawn to the experimental philosophy movement because of its naturalist appeal, should prefer naturalistic experimental philosophy over non-naturalistic one. The aim of this paper was to clarify the differences between naturalistic and non-naturalistic experimental philosophy and to outline what genuinely naturalistic experimental philosophy would look like.

## References

- Alexander, J., & Weinberg, J. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass*, 2(1), 56–80.
- Archibald SJ, Mateer CA, Kerns KA. 2001 Utilization behavior: clinical manifestations and neurological mechanisms. *Neuropsychology Review* 11: 117-130.
- Bach, Kent (1978), ‘A representational theory of action’, *Philosophical Studies* 34: 361-379.
- Brand, Myles (1979), The fundamental question of action theory. *Nous* 13: 131-151.
- Brand, Myles (1984), *Intending and Acting*. Cambridge, MA: The MIT Press.
- Cappelen, H. 2012 *Philosophy without Intuitions*. Oxford: Oxford University Press.
- Dennett, Daniel C. 1996. Seeing Is Believing: Or Is It? In *Perception*, ed. K. Akins, 158–172. Oxford: Oxford University Press.
- Dennett, Daniel C. 2005 *Sweet Dreams*. Cambridge, MA: MIT Press.
- Deutsch, M. 2010 Intuitions, counterexamples and experimental philosophy. *Review of Philosophy and Psychology* 1: 447-460.
- Darwin, Charles (1899), *The Expression of Emotion in Man and Animals*. New York: Appleton.
- Della Sala S, Marchetti C, Spinnler H. 1991 Right-sided anarchic (alien) hand: a longitudinal study. *Neuropsychologia* 29:1113-1127
- Della Sala S, Marchetti C, Spinnler H. 1994 The anarchic hand: a fronto-mesial sign. In: Boller G, Grafman J eds *Handbook of Neuropsychology, vol. 9*. Amsterdam; Elsevier, pp. 233-255.
- Fischer, E. 2014: Intuitions’ Sources. The Case of Visual Perception’s Misunderstood Phenomenology, this volume
- Frith CD, Blakemore SJ, Wolpert DM. 2000 Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 355: 1771-1788.
- Giovannetti T, Buxbaum LJ, Biran I, Chatterjee A. 2005 Reduced endogenous control in alien hand syndrome: evidence from naturalistic action. *Neuropsychologia* 43: 75-88.
- Goldman, A. (2007). Philosophical intuitions: Their target, their source, and their epistemic status. *Grazer Philosophische Studien*, 74, 1–25.
- Haggard, P. 2005a Conscious intention and motor cognition. *Trends in Cognitive Sciences* 9: 290-295.
- Haggard, P. 2005b Conscious intention and the sense of agency. In *Disorders of Volition* (Sebanz, N., ed.). Oxford: Oxford University Press
- Haidt, J. & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology*, 26, 201-218.
- Hindriks, F. 2008 Intentional action and the praise-blame asymmetry. *Philosophical Quarterly* 58: 630–41.
- Hitchcock, C. & Knobe, J. (2009) Cause and norm. *Journal of Philosophy* 106(11):587–612.

- Hornsby, Jennifer 2004 Agency and alienation. In: *Naturalism in Question*, ed. Mario De Caro & David Macarthur. Cambridge, MA: Harvard University Press.
- Israel, David, Perry, John and Tutiya, Syun 1993 Executions, Motivations and Accomplishments. *Philosophical Review* 102: 515-540.
- Jacob, Pierre - Jeannerod, Marc (2003), *Ways of Seeing. The Scope and Limits of Visual Cognition*. Oxford: Oxford University Press.
- Jeannerod, M. (1994), 'The representing brain: Neural correlates of motor intention and imagery', *Behavioral and Brain Sciences* 17: 187-245.
- Jeannerod, M. (1997), *The Cognitive Neuroscience of Action*. Oxford: Blackwell.
- Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, 10, 95–118.
- Kritikos A, Breen N, Mattingley JB. 2005 Anarchic hand syndrome: bimanual coordination and sensitivity to irrelevant information in unimanual reaches. *Brain Res Cogn Brain Res* 24: 634-647.
- Knobe, J. 2003. "Intentional Action and Side Effects in Ordinary Language." *Analysis*, 63 (2003): 190-193.
- Knobe, J. 2007. "Reason explanation in folk psychology." *Midwest Studies in Philosophy* 31 (2007): 90-107.
- Knobe, J. 2010 Person as scientist, person as moralist. *Behavioral and Brain Sciences* 33: 315-329.
- Knobe, J. & Fraser, B. (2008) Causal judgment and moral judgment: Two experiments. In: *Moral psychology, vol. 2*, ed. W. Sinnott-Armstrong, pp. 441–8. Cambridge, MA: MIT Press.
- Kornblith, H. 2014: Naturalistic Defenses of Intuition, this volume.
- Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes. *Brain* 106: 237–255.
- Liao, S. M. 2008 A defense of intuitions. *Philosophical Studies* 140:247–262.
- Lynch, M. P. (2006). Trusting intuitions. In P. Greenough & M. P. Lynch (Eds.), *Truth and realism* (pp. 227–238). Oxford: Clarendon Press.
- Machery, Edouard 2008. "The folk concept of intentional action: Philosophical and experiential issues." *Mind & Language* 23 (2008): 165-189.
- Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). Semantics cross-cultural style. *Cognition*, 92, B1–B12.
- Mallon, Ron 2008. "Knobe vs. Machery: Testing the trade-off hypothesis." *Mind & Language* 23 (2008): 247-255.
- Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical Perspectives*, IX, 185-200.
- Millikan, Ruth (2004), *Varieties of Meaning*. Cambridge, MA: The MIT Press.
- Nanay, Bence 2010a Attention and perceptual content. *Analysis* 70: 263-270.

- Nanay, Bence 2010b Morality of modality? What does the attribution of intentionality depend on? *Canadian Journal of Philosophy* 40: 28-40.
- Nanay, Bence 2010c Neither scientists, nor moralists: We are counterfactually reasoning animals. *Behavioral and Brain Sciences* 33: 347-348.
- Nanay, Bence 2011 Do we see apples as edible? *Pacific Philosophical Quarterly* 92: 305-322.
- Nanay, Bence 2012 Action-oriented perception. *European Journal of Philosophy*
- Nanay, Bence 2013 *Between Perception and Action*. Oxford: Oxford University Press.
- Nanay, Bence forthcoming Naturalizing action theory. In: M. Sprevak and J. Kallestrup (eds.): *New Waves in the Philosophy of Mind*. Palgrave Macmillan.
- Nichols, Shaun & Joseph Ulatowski 2007. "Intuitions and individual differences: The Knobe effect revisited." *Mind & Language* 22 (2007): 346-365.
- Nichols, S., & Knobe, J. forthcoming Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*.
- Nichols, S., Stich, S., & Weinberg, J. (2003). Meta-skepticism: Meditations on ethno-epistemology. In S. Luper (Ed.), *The skeptics* (pp. 227–247). Aldershot, U.K: Ashgate Publishing.
- O'Shaugnessy, B. 1980 *The Will*. Cambridge: Cambridge University Press.
- Pacherie, Elisabeth 2007 The anarchic hand syndrome and utilization behavior: a window onto agentic self-awareness. *Functional Neurology* 22: 211-217.
- Perry, John 2001 *Knowledge, Possibility and Consciousness*. Cambridge, MA: The MIT Press.
- Petrinovich, Lewis, and O'Neill, Patricia. 1996. "Influence of Wording and Framing Effects on Moral Intuitions." *Ethology and Sociobiology* 17, 145-171.
- Prinz, Jesse, J. 2008 Empirical philosophy and experimental philosophy. In: All consciousness is perceptual. In: J. Knobe and S. Nichols (eds.): *Experimental Philosophy*. New York: Oxford University Press.
- Pust, J. (2000). *Intuitions as evidence*. New York: Garland.
- Quine, W.V.O. (1969) Epistemology naturalized. In: Quine: *Ontological Relativity and Other Essays*, New York: Columbia University Press.
- Quine, W.V.O. (1984) Reply to Putnam. In: L. E. Hahn and P. A. Schilpp (eds): *The Philosophy of W. V. Quine*. La Salle: Open Court.
- Schnall, Simone, Jennifer Benton and Sophie Harvey 2008 With a Clean Conscience. Cleanliness Reduces the Severity of Moral Judgments. *Psychological Science* 19: 1219-1222.
- Searle, John (1983), *Intentionality*. Cambridge: Cambridge University Press.
- Shallice, T. and Burgess, P. W. 1991 Deficits in strategy application following frontal lobe damage in man. *Brain* 114: 727-741.

- Shallice, T., Burgess, P. W., Schon, F., and Baxter, D. M. (1989). The origins of utilization behaviour. *Brain* 112: 1587–1598.
- Simmons, Daniel J. - Chabris, Christopher F. 1999 Gorillas in our Midst: sustained inattentive blindness for dynamic events. *Perception* 28: 1059-1074.
- Sinnott-Armstrong, W. (2006). Moral intuitionism meets empirical psychology. In T. Horgan & M. Timmons (Eds.), *Metaethics after Moore* (pp. 339–365). New York: Oxford University Press.
- Slezak, Peter 1987 Intending and acting (book review). *Journal of Philosophy* 84: 49-54.
- Slezak, Peter 1989 How NOT to naturalize the theory of action, In: Peter Selzak and W. R. Arbury (eds.): *Computers, Brains and Minds*. Dordrecht: Kluwer, pp. 137-166.
- Sosa, E. (2005). A defense of the use of intuitions in philosophy. In M. Bishop & D. Murphy (Eds.), *Stich and his critics*. Blackwell.
- Sosa, E. (2006). Intuitions and truth. In P. Greenough & M. P. Lynch (Eds.), *Truth and realism* (pp. 208–226). Oxford: Clarendon Press.
- Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, 132, 99–107.
- Stich, S. forthcoming Experimental philosophy and the bankruptcy of the ‘Great Tradition’.
- Swain, S., Alexander, J., & Weinberg, J. forthcoming The instability of philosophical intuitions: Running hot & cold on truetemp. *Philosophy and Phenomenological Research*,
- Tversky, Amos, and Kahneman, Daniel. 1981. “The Framing of Decisions and the Psychology of Choice”. *Science*, 211, 453-458.
- Valdesolo, P. & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, 18, 689-690.
- Velleman, David 2000 *The Possibility of Practical Reason*, Oxford: Oxford University Press.
- Weinberg, J., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, 29(1&2), 429–460.
- Williamson, T. (2004). Philosophical ‘Intuitions’ and skepticism about judgment. *Dialectica*, 58, 109–153, p. 114.