



The feeling of agency hypothesis: a critique

Grünbaum, Thor

Published in:
Synthese

DOI:
[10.1007/s11229-015-0704-6](https://doi.org/10.1007/s11229-015-0704-6)

Publication date:
2015

Document version
Early version, also known as pre-print

Citation for published version (APA):
Grünbaum, T. (2015). The feeling of agency hypothesis: a critique. *Synthese*, 192(10), 3313-3337.
<https://doi.org/10.1007/s11229-015-0704-6>

The Feeling of Agency Hypothesis: A Critique

By Thor Grünbaum (Philosophy, University of Copenhagen)

Abstract

A dominant view in contemporary cognitive neuroscience is that low-level, comparator-based mechanisms of motor control produce a distinctive experience often called the feeling of agency (the FoA-hypothesis). An opposing view is that comparator-based motor control is largely non-conscious and not associated with any particular type of distinctive phenomenology (the simple hypothesis). In this paper, I critically evaluate the nature of the empirical evidence researchers commonly take to support FoA-hypothesis. The aim of this paper is not only to scrutinize the FoA-hypothesis and data supposed to support it; it is equally to argue that experimentalists supporting the FoA-hypothesis fail to establish that the experimental outcomes are more probable given the FoA-hypothesis than given the simpler hypothesis.

1. Introduction

A number of prominent cognitive neuroscientists and philosophers have argued that a distinctive feeling of agency plays an important role in individual motor cognition. For instance, Haggard and Chambon (2012) write in a recent paper: “Most of us have the feeling that we are in control of what we are doing most of the time: this is the normal sense of agency” (R390). In a discussion of possible computational mechanisms producing the feeling of agency, they add, “The prospective aspect of agency gives a clue as to *why* the brain might provide a *distinctive conscious experience of being in control*” (R392, my emphasis).¹ And Bayne and Pacherie (2007) write: “agentive *experience* – that is, our moment-by-moment

¹ For similar statements, see Chambon and Haggard, 2013, 359, and Kühn and colleagues, 2013, 1936: “We can, and frequently do, make instrumental actions where we have a definite background feeling or buzz of being in control.”

sense of ourselves as the agents of various movements – is largely the output of low-level, comparator-based systems” (485). Elsewhere in their paper, they elaborate: “A strong case can be made for thinking that the contents of agentic experience can go beyond mere representation of oneself as the agent of an event, but also include information about degree of control one has over the movement and the degree to which the action is effortful “(477).

These authors endorse the idea that when an agent executes a voluntary movement, then she will normally experience a distinctive feeling of agency that should be explained as the output of comparator mechanisms involved in online motor control. This feeling of agency is thought to be distinctive in two ways: First, it is claimed to have a distinctive type of phenomenology, and, second, it is claimed to have a distinctive type of content representing one’s movements as self-caused (and maybe the degree of self-causing). For the sake of simplicity, I will call the view that voluntary movements are accompanied by a distinctive feeling of agency that is the low-level output of comparator-based systems of online motor control the *Feeling of Agency-Hypothesis* (FoA-hypothesis).

No one should doubt that we have experiences when we perform actions. In an ordinary situation, when a person intentionally opens a door, she will consciously intend to open it, have proprioceptive and tactile experiences as she moves, grasps, turn, and pull the doorknob, have visual experiences of the door and her own body and their relative movements; she will furthermore automatically form low-level sensory anticipations of the most likely sensory consequences of her movements, as well as higher-level rational expectations of action-outcomes. In addition, there might be experiences such as some feeling of effort and fatigue and various affective experiences. Some of these experiences or constellations of experiences might be distinctive of intentional action, though most likely not in the sense of being necessary and sufficient for intentional action.

Accepting that the performance of intentional action normally is accompanied by a set of cognitive, bodily, and visual experiences is not the same as accepting that there is a special feeling of agency produced by low-level comparator mechanisms of motor control. One can accept that there are conscious mental states or processes that are proprietary to action, such as practical deliberation, practical decision-making and intending, without accepting that there is a special phenomenal feeling associated with motor control. And one can do this while still accepting that voluntary movements are enabled by low-level, comparator-based systems for motor control. For these reasons, it becomes important to evaluate the empirical justification for claiming that the normal operation of comparator-based mechanisms of control of voluntary movement produces a distinctive feeling of agency.

The target of this paper is a prominent scientific model of voluntary motor action and motor cognition according to which the performance of voluntary movements is associated with a low-level, comparator-based, phenomenologically distinctive feeling of agency – in addition to the conscious states and experiences listed above.² In section 2, I outline in more detail the FoA-hypothesis. Particular emphasis is put on the supposed rational and explanatory role played by the distinctive feeling of agency. In section 3, I lay out the structure of my argument. The prior probability of the FoA-hypothesis is not higher than the prior probability of a simpler hypothesis according to which low-level, comparator-based systems for online motor control do not produce any distinctive agentive experience. It therefore becomes important to evaluate whether data that are often assumed to support the FoA-hypothesis really are more probable given this hypothesis than given the simpler hypothesis. Consequently, in Section 4, I look at data from action-recognition paradigms, sometimes assumed to be best explained and predicted by the FoA-hypothesis, and, in section

² One referee wrote that the argument presented in this paper generalizes to other models of the phenomenal sense of agency, such as optimal cue integration (Synofzik, Vosgerau, and Newen, 2008) and some versions of Wegner's inferential model (e.g. Wegner, 2004). Another referee insisted, however, that my arguments only have a limited scope and leave other models untouched. I will leave these issues of scope up to the reader.

5, I consider a perhaps more direct form of evidence for the FoA-hypothesis, namely, evidence from experimental paradigms using introspective reports of degrees of one's feeling of agency. In all cases, a simpler explanation – one that does not include a distinctive feeling of agency – is available. If the experimental data do not support the FoA-hypothesis better than the simpler hypothesis, we must look elsewhere for a reason to believe that the FoA-hypothesis is the most probable hypothesis. In section 6, I then consider and reject two non-experimental, phenomenological arguments in favour of the FoA-hypothesis. Finally, I sum up and conclude in section 7.

2. The Feeling of Agency-Hypothesis

The FoA-hypothesis is the conjunction of four claims:

1. A person's ability to make voluntary movements is best explained by a dedicated, modular comparator-based control system.
2. Normally, activation of the comparator-based control system during voluntary movement produces as output a distinctive feeling of agency.
3. The feeling of agency has a distinctive type of phenomenal feel or quality.
4. The feeling of agency has a distinctive type of representational content.

Before discussing these claims, let me briefly mention some other features often associated with the comparator-based feeling of agency. Without committing myself to any view of non-conceptual content, a number of properties often associated with non-conceptual content follow from the four claims. If the (1) and (2) are correct, then the feeling of agency is the output of a modularized control mechanism. This indicates that the feeling of agency is belief-independent in the sense that if the comparator-based mechanisms is activated in the right way, it will output a feeling of agency independently of the agent's beliefs about her

action and its context (Bayne and Pacherie, 2007). It also seems to follow that the feeling of agency is not the product of a personal level inference and is not a judgement.³

Let me outline a number of considerations in favour of the four claims before advancing my criticism. Claim 1 is that control of voluntary movements is best explained by a dedicated, modular comparator-based control system. The basic idea is well known. For the sake of the argument of this paper, we can assume that Claim 1 is true.⁴ Therefore, it will not feature prominently in the ensuing discussion.

Claim 2 states that, normally, activation of the comparator-based control system during voluntary movement produces a distinctive feeling of agency. According to one proposal about how such a motor control mechanism could produce a feeling of agency, the feeling is the output of a comparator successfully matching predicted sensory consequences of motor execution (computed by the so-called forward model) with sensory feedback from the execution (Bayne and Pacherie, 2007). According to another proposal, the feeling is the output of a successful matching between forward model prediction and the intended motor goal (Gallagher, 2000). A third proposal is that the feeling is caused by the so-called inverse model computing the motor commands needed for successful execution of the motor goal (Haggard & Johnson, 2003). Finally, according to a fourth proposal, the low-level feeling of agency is jointly caused by signals from action-selection in the inverse model and signals from the comparator matching predicted sensations with feedback from the execution (Haggard & Chambon, 2012). Many important and decisive questions remain with respect to each of these proposals and their relation to each other. For the remainder of this paper, I will, however, put these issues of detail to the side. My critical discussion of the FoA-hypothesis is aimed at all versions of the comparator-based account of the feeling of agency. By the phrase

³ See also Haggard and Tsakiris, 2009, 243, and Wenke et al., 2010, 35.

⁴ It should be noted that the “classical” comparator model of motor control is not unchallenged. See, for example, Friston, 2011.

“comparator-based control system” I refer to the general idea that modular and dedicated mechanisms involved in online motor control in addition to their function in motor control also produce a distinctive, low-level feeling of agency.

One possible argument for the comparator-based FoA-hypothesis focus on a functional role of the comparator-based identification of a movement as self-produced. The need for some such mechanism arises from the fact that an agent’s body is moving all the time. Some of these movements will be internally caused by the agent’s intentions, whereas countless others are caused by external forces. It is an important task for the motor system to keep track of this difference between internally caused and externally caused movements. One proposal is that an important mechanism for keeping track of the causal origin of one’s movement is a mechanism for comparing sensory feedback from the execution of the movement with a prediction of the sensory feedback. To quote Jeannerod: “simulation of the outcome of an action by the internal model can be a powerful means for action identification: if and only if the reafferent signals match the expectation of the internal model, can the action be definitely identified as self-generated” (Jeannerod 2009, 528). If there is a match between prediction and feedback, the movement is self-generated and will be tagged with a feeling of agency. Without such a tagging mechanism, the agent would not be able to know which of her movements are her actions (see Farrer et al., 2008, Hohwy, 2007, Sato and Yasuda 2005). So, according to this view, the mechanisms responsible for low-level, comparator-based action-identification are also responsible for generating the feeling of agency.⁵ It would seem like a small step from this claim to assume that the content of the feeling of agency is representing one’s movement as self-generated – a distinctive representational content. Consequently, describing the feeling of agency as the output of low-level, comparator-based

⁵ Similar functional stories could be told with respect to the other comparator-based accounts of the feeling of agency.

action-identification also provides us with one way to understand the “distinctiveness” (claim 4).

In conceiving of the feeling of agency as a distinctive conscious output of a particular kind of low-level, modular comparator mechanism, the feeling of agency is distinguished from non-motor, high-level cognitive phenomena like practical decision-making and intentions.⁶ There are a number of related reasons for insisting on such a dissociation between conscious intention and the feeling of agency. First, one might want to distance one’s model from inferential post hoc models of sense of agency, such as Wegner’s (Wegner and Wheatley, 1999), by tying the feeling of agency closely to mechanisms of online motor control or action-selection (for such reasons, see, e.g., Moore and Haggard, 2008). Second, one might want to say that habitual actions, which are not preceded or accompanied by any conscious intention, are still associated with a feeling of agency (see, e.g., Bayne and Levy, 2006). Third, whereas the feeling of agency is the product of low-level, modular comparator mechanisms and as such is cognitively impenetrable, intentions have propositional content and are the output of inferential activity (see, e.g., Bayne and Pacherie, 2007). Fourth, a final reason for distinguishing between, on the one hand, the feeling of agency and, on the other hand, reasoning and intentions is the role played by the assumed “intention-free” feeling in the explanation of experiences associated with certain neuropsychological and psychopathological conditions. I will return to this issue in a Section 6.

In sum, the feeling of agency is assumed to be a distinctive mental component in addition to the mental components already acknowledged by standard theories of action (conscious intention, conscious decision, proprioceptive awareness, perception of consequences), and not reducible to any of these elements or their combination. On the FoA-

⁶ By “intention” I mean a personal-level state, which is often the outcome of the agent’s practical decision-making. It is a state governed by particular normative and rational constraints. As such, this state is very different from the kind of functional state sometimes referred to by the term “motor intention”. For discussion of this point, see Butterfill & Sinigaglia, 2014.

hypothesis, the feeling of agency thus conveys “sparse intention-free information about the ways in which the agent is moving” (Bayne and Pacherie, 2007, 488).

This distinctiveness makes the feeling of agency a strong candidate for a defeasible mark of voluntary action.⁷ This means that, most of the time, our voluntary movements are associated with a feeling of agency, and if the feeling is missing or is diminished in some way, this would be a reason for judging that one’s movement is caused by some external agent. Let us call this type of judgement in which a person judges that some event (be it a movement or some external environmental event) is caused by herself *agency judgements*. If the feeling of agency is a distinctive conscious mark of voluntary action, then feelings of agency would seem to be ideally suited to epistemically ground agency judgements. Many researchers thus assume that, in the normal case, an individual’s feeling of agency for some event is her epistemic ground for judging that she is the cause of the event. Assuming this type of empiricist model, Haggard and Tsakiris (2009, 243) claim: “Under normal circumstances, the FoA [feeling of agency] is a necessary condition for JoA [judgement of agency], and indeed forms the evidence base for the judgment.”⁸

To sum up, according to the FoA-hypothesis, when an agent performs a voluntary movement, low-level comparator-based mechanisms of motor control produce a distinctive feeling of agency. In virtue of its distinctiveness, the feeling of agency functions as a subjective mark of agency and as such serves as evidence base for agency judgements.⁹

⁷ See, e.g., Tsakiris and Haggard, 2005, 387.

⁸ See also Bayne and Pacherie (2007): “agentive judgments are typically grounded in and justified by agentive experiences. In the normal case, we judge that we are the agent of a particular movement on the grounds that we enjoy an agentive experience with respect to it; here, our agentive judgments are simply endorsements of our agentive experiences” (477).

⁹ How does my use of the term “feeling of agency” relate to the more frequent use of the term “sense of agency” that we see in the literature? It is hard to give a clear answer because of the often ambiguous and loose terminology. Often the two terms appear to be simply synonymous but at other times not. According to one prominent way to use “sense of agency”, it refers to a more complex phenomenon involving both low-level phenomenal states and high-level cognitive evaluations and judgements (see Synofzik, Vosgerau, & Newen, 2008, and Pacherie, 2008, and Gallagher, 2012, for comparable complex models). In order to distinguish the target of my critique from this type of more global, complex model, I will stick to the term “feeling of agency”. On some versions of the complex model, something like the FoA-hypothesis is an element.

3. A Simpler Hypothesis

In this paper, I pursue the question of whether experimental data from agency-paradigms in cognitive neuroscience support the FoA-hypothesis. I do this in the spirit of contrastive hypothesis testing according to which a given hypothesis is tested by asking whether the data is more likely on the hypothesis in question than on some other hypothesis (Sober, 1999, 2008, Machery, 2013). In the present context, the relevant contrasting hypothesis is a simpler hypothesis that posits all the same elements as the FoA-hypothesis except for the feeling of agency (thus rejecting claims 2-4). This means that the simpler hypothesis explains agency judgements not by grounding them in feelings of agency but instead in intentions, perceptual expectations, and perceptual experience. I will thus pursue the question of whether data from agency-paradigms support the FoA-hypothesis by asking whether data from relevant agency-paradigms are more probable given the FoA-hypothesis than given the simpler hypothesis. If both hypotheses do an equally good job in predicting the data, then if there are no substantial theoretical or common sense intuitions in favour of the FoA-hypothesis, we should prefer the simpler hypothesis. That is, we should stop mentioning the feeling of agency in our explanations of voluntary action and associated motor cognition.

Let me say more about the simpler hypothesis and my “hypothesis-testing” strategy. A central motivation for the FoA-hypothesis as an empirical hypothesis about the mechanisms of motor cognition is its ability to explain the assumed phenomenology of voluntary movement as well as an agent’s ability to judge that some environmental event is her action or a consequence of her action. In a way, it is an elegant hypothesis as far as it explains the ability to judge by explaining the phenomenology. The FoA-hypothesis explains agency judgements by the feeling of agency and explains the feeling of agency as the output of the comparator-based control mechanisms. But this simplicity is only apparent. The FoA-

hypothesis is not denying the existence of practical reasoning, intentions, proprioceptive experiences, and visual experiences. These additional elements are part of the mental context of acting and are important factors in everyday motor cognition.¹⁰

The simpler hypothesis explains the agent's ability to make agency judgements by these additional conscious states, the existence of which is acknowledge also by proponents of the FoA-hypothesis. When an agent judges whether some event is her own action, she draws on various forms of information from her intention, bodily awareness, and vision. The simpler hypothesis is simpler because it assumes all the same mental elements as the FoA-hypothesis except for the distinctive feeling of agency as the output of comparator-based motor-control processes. For now, this is all I will say about the simpler hypothesis. To be sure, important questions remain. Such as, is there any role left for the comparator mechanisms to play in motor cognition according to the simpler hypothesis? And what is the role of perceptual expectations? I will briefly return to these questions in Section 7.

This should suffice to get the argument going. For now, let us assume that prior to or independent of experimentally based observations, the simpler hypothesis (according to which agency judgements are explained by intentions, perceptual expectations, and perceptual feedback) seems just as likely as the FoA-hypothesis (according to which agency judgements are explained by the feeling of agency). The question is whether experimental data tip the evidential balance in favour of the FoA-hypothesis. It therefore becomes important to evaluate whether the FoA-hypothesis is better supported by experimental data than the simpler hypothesis. In the following sections, I will therefore ask whether data from relevant experimental paradigms are good evidence for the FoA-hypothesis. The general thrust of the argument is to compare informally the likelihoods of the two competing hypothesis: the probability of the experimental data given the FoA-hypothesis compared to

¹⁰ As has been thoroughly documented by Jeannerod (1997, 2006)

the probability of the experimental data given the simpler hypothesis. I will assume that, given the equal prior probability of the two competing hypotheses, if the two likelihoods are the same, then the data is uninformative with respect to our choice of hypothesis. If the data is equally probable given both hypotheses ($p(\text{data} \mid \text{FoA-hypothesis}) = p(\text{data} \mid \text{simpler hypothesis})$), then the data itself cannot serve as evidence for one theory over another. If the latter turns out to be the case, then, in order to argue for the FoA-hypothesis, one would have to look elsewhere than contemporary experimental science.

It could be objected that the simpler hypothesis is committed to a simplistic epistemic bipartition of movements into experience of intentional action and experience of involuntary movements, whereas the FoA-hypothesis is motivated by an epistemic tri-partition of movements into experience of intentional action, experience of voluntary movements, and experience of involuntary movements. That is, the FoA-hypothesis is motivated by the idea that some movements are experienced as voluntary even if they are not performed with a conscious intention. That is, it is motivated by the idea that sometimes agents can know they are acting even if they do not know what their intentions are.¹¹ According to the objection, the simpler hypothesis cannot explain the experience-based distinctions between, on the one hand, intentional action and voluntary movement, and, on the other hand, voluntary and involuntary movement. Since agents can report a difference between voluntary and involuntary movements in situations where they are not acting intentionally, their reports must be grounded on their feelings of agency and not their intentions. My criticism thus seems to overlook or misconstrue the real explanatory target of the FoA-hypothesis – a target which is out of reach for simpler hypothesis.

My reply to this objection is threefold. First, it is wrong to think that the simpler view is inconsistent with a tri-partition view of action. The simpler view is inconsistent only with

¹¹ I thank a reviewer for providing me with this sentence.

the view that the difference between voluntary and involuntary action implies a difference in presence and absence of a feeling of agency.

Second, I reject the claim that we have any clear examples of reports of voluntary agency in cases where the agent is not performing an intentional action. Such cases would have to be instances where no intentions are present. Let us briefly look at some of the categories of cases often mentioned.

Habitual action: Buckling the belt when getting into one's car or the routine involved in brushing one's teeth in the morning might not involve any active reasoning and decision-making. Be that as it may, it is still very plausible that these actions are controlled by long-term intentions (policies) and barely conscious practical reasoning and intentions required to navigate trivial problems in a dynamic world. I see no reason why we should think that actions like reaching for my coffee as I work at my computer do not involve basic forms of practical reasoning and context-dependent intentions.

Sub-intentional action: Mindlessly tapping one's finger during a lecture or picking one's nose in public are voluntary actions but plausibly not intentional – after all, the agent would not have consciously decided to perform these actions in these social contexts. These actions are voluntary because, on becoming aware of them, the agent has the power to inhibit them. This much I accept. What I deny is that before becoming aware of them, the agent is experiencing a feeling of agency. One possible account of these situations would be that on becoming aware of these actions, the agent becomes aware of actions that she can either endorse (i.e. perform intentionally) or inhibit (intentionally stop performing).

Autonomic actions: Blinking and breathing are plausible cases of actions which are voluntary but not intended by the agent. They are, however, also cases of which the agent is most of the time not conscious and where performance does not require any consciousness.

Explaining the difference between an involuntary movement and an autonomic action does not require any mention of the feeling of agency.

To sum up, these are cases that could at most establish a metaphysical tri-partition of movements (intentional movement, voluntary movement, and involuntary movement) but not a corresponding epistemic tri-partition in experiences of movement requiring us to accept an intention-free feeling of agency.

Third, this paper is primarily concerned with experimental arguments in favour of the FoA-hypothesis. There are to my knowledge no experimental paradigms that support the view that the difference between voluntary and involuntary movement involves a distinctive, intention-free feeling of agency. The simple reason for this fact is the following. Experimental action-paradigms require participants to perform tasks as instructed by the experimenter. Participants are thus always executing intentional actions. So, even if we can make theoretical sense of a metaphysical and epistemic tri-partitions of action, it will in principle be extremely difficult to design experiments to investigate it. When discussing the experimental designs, we are therefore justified in employing a simple bipartition of action into intentional action and involuntary movements.

4. Evidence from Action-Recognition Paradigms

By assuming an epistemic relation between feelings of agency and agency judgements, the proponents of the FoA-hypothesis are provided with a way to interpret data from action-recognition studies. A typical design is one in which participants are making a manual action while watching a monitor showing either their own hand's moving or someone else's. The task is then to judge whether they are seeing their own bodily movement (Daprati et al. 1997, Daprati et al. 2007, Farer et al. 2003, van den Bos and Jeannerod 2002). Other studies have extended this paradigm from producing and watching manual action to judgements about

who produced (by movements of a joystick or key presses) an arbitrary effect, such as a tone or movement of an object on a monitor (Ritterband-Rosenbaum et al. 2011, Sato and Yasuda 2005, Sato 2009, Wegner and Wheatley 1999). If we assume that participants' agency judgements are grounded in their feelings of agency, then we can interpret their judgments as a measure of their feeling of agency.¹² If we can systematically manipulate participants' agency judgements, then we would most likely be doing it by manipulating their feeling of agency, or so the assumption would be.

This assumption seems to receive additional support from the experimental data showing that predictive mechanisms play an important role in making agency judgements. The closer the match between predictions (based on action-effect associations learned in a training phase of the experiment) and sensory events resulting from a person's movements, the higher the probability that the person judges the event to be under her control (e.g. Sato and Yasuda, 2005). Such data are consistent with the idea that forward models compute sensory predictions that are subsequently matched with sensory feedback from the movement by a comparator mechanism. If we assume that the output of the comparator is a feeling of agency, then it seems reasonable to claim that in manipulating predictions and their match with feedback, we are also manipulating the feeling of agency. Further evidence for the assumption that we can measure the feeling of agency by collecting agency judgements comes from data by Tsakiris and colleagues (2005) suggesting that the decisive factor in making a correct agency judgement is efferent signals that the authors relate to forward models and sense of agency.

¹² See Farrer et al. (2003): "[the feeling of agency] is the feeling that leads us to attribute an action to ourselves rather than to another person (324)".

Given these considerations, the question now becomes whether we nevertheless have good reasons for thinking that the likelihoods of the FoA-hypothesis and the simpler hypothesis are the same.¹³

As a paradigmatic example of an action-recognition task, let us look at the seminal 1997-study by Daprati and colleagues. In this study, the task was to perform a requested movement and monitor its execution by looking at an image of the movement in a mirror. Once the movement was performed, the image would become dark and a question was asked to the subject: “You have just seen the image of a moving hand. Was it your own hand? Answer YES if you saw your own hand performing the movement you have been executing. Answer NO in any other case, that is if you doubt that it was your own hand or your own movement.” Daprati et al. tested two sets of movements: extension of a finger or movement of a joystick by wrist rotation. For each set of movements, one of three possible images of the hand could be presented: participants own hand, experimenter’s hand performing the same movement, or experimenter’s hand performing a different movement. The data showed that participants made virtually no errors of judgement when watching their own movement or the experimenter’s hand performing a different movement. By contrast, errors were made when watching the experimenter’s hand performing the same movement. Data showed that “[normal participants] misjudged the hand as theirs in about 30% of cases” (Daprati et al., 1997, 81). Interestingly, data also showed that schizophrenic participants suffering from hallucinations and/or delusions of alien control made more frequent errors in this condition (Experimenter-Same) in comparison to normal controls (hallucinating 77%, hallucinating + delusions 80%, delusions 50%).

Between conditions, type of voluntary movement and whose hand is being projected are being manipulated; but across all conditions, the fact that the agent is performing an

¹³ $p(\text{data} | \text{FoA-hypothesis}) = p(\text{data} | \text{simpler hypothesis})$

intentional action remains constant. At no point is there any reason to think that the subject is in doubt about whether she is acting or in doubt about what she is doing. She has been asked to move her hand and she complies with the instruction. It is important to notice that the doubt manifested by participants in the crucial experimental condition (experimenter's hand, same movement) concerns not the participant's own action but the visual image of a hand's movement in the mirror. The participant is not in doubt about what she is doing with her hand underneath the occluded surface but is in doubt about whose action she is seeing on the monitor. Thus, in all conditions, the subject knows which movement she is supposed to perform and which movement she is in fact performing, she has proprioceptive feedback from her moving hand, and she has visual information from the image. This much ought to be acceptable to both the FoA-hypothesis and the simpler hypothesis. The FoA-hypothesis is distinguished by its postulation of an extra element in order to explain the agency judgements.

This addition is not needed. The simpler hypothesis is that forms of information (intention, proprioception, vision), accepted by both hypotheses, provide the subject with all she needs in order to make agency judgements in this recognition task. Given these forms of information, it is to be expected that the more similar the movement of the experimenter's hand is to the movement performed by the participant, the more mistakes she will make in judging whether she is looking at an image of her own hand. When I move my hand in a well-rehearsed way, I expect it to feel and look a certain way. If, when I am moving, I am looking at an image of a hand's movement that has a sufficient number of the expected kinematic properties, it is only to be expected that I have a tendency to misjudge that the hand is my own. This interpretation is supported by data from Daprati et al. (2007) demonstrating that visual perception of the kinematic properties of the image can be sufficient to explain

action-recognition. Hours after performing a movement, participants can recognise an image of their own movement in virtue of its kinematic features.

To be sure, it might be the case that the best explanation of the subject's ability to match her knowledge of her motor goal, proprioceptive awareness, and visual anticipations to the kinematics of the movement represented in the mirror image involves underlying comparator-based systems. But this provides us with no independent reason for adding the feeling of agency to the picture. For as Jeannerod remarks about the low-level comparator-based form of identification of movements as self-caused: "in everyday life, action identification appears to be a largely automatic process. Subjective awareness does not seem to be involved (2009, 528)".

It could be objected that this interpretation of the experiment misunderstands the motivation for the FoA-hypothesis and the role of the 1997-Daprati experiment in the argument for the hypothesis. One important reason for endorsing the FoA-hypothesis is a rejection of a standard model of knowledge of action according to which an agent is aware of her action by being aware of her intention and her bodily and perceptual experiences while acting. One argument for this rejection is the claim that schizophrenic agents suffering from delusions of alien control can deny agency for their own actions even though they are aware of their intentions and perceptual experiences. Intentions and perception are therefore insufficient to explain action awareness and ground knowledge of action. Something is missing. I will return to this style of argument in section 6. For now, the important point is the claim that it is in the context of the rejection of the standard model of agents' knowledge that we should to understand the 1997-Daprati results: the increased error rate by schizophrenic participants must be explained by something else than intentions and perception, namely, the diminished or altered feeling of agency.

This inference is highly problematic. The rejection of the standard model of knowledge of action is not convincing. It might be correct that for a schizophrenic agent suffering from delusions of alien control, awareness of her intention and perceptual experiences are insufficient for knowing her own action (she is aware of her intentions and perceives the consequences of the execution but believes some external agent is controlling her movements).¹⁴ This fact has, however, no immediate implications for the normal case. Plausibly, delusions of alien control involve odd or strange experiences when acting.¹⁵ These experiences would act as defeaters. One explanation of the poorer performance of patients experiencing alien control in action-recognition tasks could therefore be that they have problems in matching their intentions and performance due to certain movement related error signals.¹⁶ In the normal case and in the absence of such defeating “error-signals”, awareness of intention and perceptual experiences are all the agent need to know what she is doing.

To conclude, the data from Daprati et al.’s action-recognition task tell us nothing about the normal feeling of agency. The same argument applies to the explanation of behaviour in other versions of the action-recognition paradigm. The behavioural data do not compel us to add the feeling of agency to the mix. The data therefore provide us with no reason for accepting Bayne and Pacherie’s claim that “Agents will typically judge that they are the authors of a movement – that it realizes one of their own actions – if and only if they have an agentic experience with respect to it” (2007, 486) – where by “agentic experience” the authors mean a distinctive feeling of agency produced by low-level, modularized comparator mechanisms for motor control. In sum, data from these paradigms do not compel us to accept that the FoA-hypothesis offers a better explanation and prediction than the

¹⁴ The experimental evidence for this claim is Spence et al., 1997.

¹⁵ Frith & Johnstone (2003, 137-138) describe these odd experiences as “unexpected sensations” “that indicate that we are not fully in control of our movements” and that “might feel like unpredictable passive movements”.

¹⁶ This seems to be the explanation offered by Frith & Done, 1988, and Stirling, Hellewell, & Quraishi, 1998. Notice that Daprati et al. did not find that patients with delusions of alien control were worse at action-recognition than patients experiencing hallucinations.

simpler hypothesis. The data show only that by manipulating sensori-motor predictions, we can influence agency judgements.

5. Evidence from Introspective Paradigms

The fact that the feeling of agency is not needed to explain participants' agency judgements in action-recognition tasks does not show that there is no distinctive feeling of agency. It only shows that we do not need to assume such a feeling to explain data from this type of task. The feeling of agency might still be needed to explain data from other types of task. Maybe I have not considered the right kind.

The problem could be that agency judgements in recognition tasks are too indirect to function as a measure of the feeling of agency. Maybe we need a more direct measure. Recently, a number of studies have taken up this challenge. Instead of asking participants to make agency judgements (with respect to external events), participants in these studies are asked to report their feeling of agency.¹⁷ If participants in these studies are reporting consistently and reliably a distinct feeling of agency, a plausible explanation would be that participants experience a distinct feeling of agency. Unless we have independent reasons not to trust the participants, we should trust their subjective reports. Let us, therefore, take a closer look at some of the studies that allegedly measure the subjective feeling of agency.

A word of caution is in order before looking at the studies in question. Not all of the studies belong to the framework of the FoA-hypothesis. In particular, the study by Wegner and colleagues (see below) is sometimes interpreted as going against the FoA-hypothesis in so far as it purports to show that the feeling of agency is not the output of low-level comparator mechanisms but instead the product of high-level cognitive inferences (Synofzik,

¹⁷ It is an open question whether introspective reports are best understood as a judgement about one's experience or as an expression of one's experience (see, e.g., Bar-On, 2004). This dispute has no implications for the present argument.

Vosgerau, and Newen, 2008, but see Carruthers, 2012). One common assumption, accepted by all parties to this discussion, is that irrespective of the level at which the causal mechanisms responsible for the feeling of agency are located, everyone – experimental subjects as well as researchers – is referring to the same distinctive phenomenal experience.¹⁸ It is this feeling that the participants are assumed to report when quizzed about their feeling of agency. In what follows, I will question this assumption and thereby undercut the idea that such studies can be used in support of the FoA-hypothesis. That is, I will set aside the issue of how the reported feeling fits the theoretical framework motivating a given study and simply focus on the alleged introspective reports of the feeling of agency.

Do these studies provide evidence for the claim that the performance of voluntary action is associated with a phenomenologically distinctive feeling of agency? If the answer is yes, then the FoA-hypothesis is back in business. Data from these introspection studies would then be more probable given the FoA-hypothesis than given the simpler hypothesis (since according to the latter there is no feeling of agency). I will look at two paradigmatic experiments, Wegner et al. (2004) and Wenke et al. (2010), both of which has been extended and used in other studies (see, for example, Sato, 2009, and Chambon et al., 2013). In both cases, my main argument is that there are plausible alternative interpretations of the data that make no reference to a feeling of agency. My main thrust of argument is that we have reasons for thinking that participants are interpreting the introspective task differently than the experimenters. The experimental situations are highly artificial and the “introspective questions” are very unusual, so it might not be obvious for participants how to comply with the instruction to give a rating of their feeling of control. In the following, I will in turn

¹⁸ As Wegner writes: “Although the proper experiments have not yet been done to test this, it seems likely that people could discriminate the feeling of doing from other feelings, knowing by the sheer quality of the experience just what has happened” (Wegner, 2004, 658).

briefly describe each study and some reasons for accepting an alternative interpretation of the data (in line with the simpler hypothesis) before discussing a general methodological concern

Wegner et al., 2004: In this study, participants were not moving themselves but were watching the movements of another person. Behind each participant was a “hand helper” with her arms placed where the participant’s arms would normally be. In front of the participant was a mirror in which she could see her own face and torso and the arms of the hand helper positioned in such a way as to appear to be the participant's own arms. The participants were equipped with headphones and asked to make a number of ratings of their experience of vicarious hand movements. Participants were divided in two groups. On each trial, the preview group was given verbal instructions (e.g. “wave hello with the right hand”) in the headphones just before the hand helper carried out the instruction. The no-preview group heard nothing in the headphones (other things being equal). After finishing the exercise (26 instructions or 3.5 min) and taking off all of the equipment, participants answered two questions (“How much control did you feel that you had over the arms’ movements?”; “To what degree did you feel you were consciously willing the arms to move?”) by a rating on a 7-point scale from 1 = not at all to 7 = very much. The results showed that the preview group rated their feeling of control as being higher than the no-preview group (the mean score was 3.00 compared 2.05). According to Wegner at al., the data demonstrate that the participants in the different conditions experience a feeling of agency with varying strength: strongest in the preview group, weakest in the no-preview group. If this interpretation is correct, then the data seems to be in conflict with the simple hypothesis. If participants are really experiencing a feeling of agency and expressing it in their reports, then the data (the introspective reports) are more probable given the FoA-hypothesis than given the simpler hypothesis according to which there is no feeling of agency.

Alternative interpretation: In the Wegner et al. study, in the experimental conditions, participants are fully aware they aren't performing the action they see – in fact, in nearly all conditions, participants are fully aware of the fact that they are not performing any actions (Wegner et al., 2004, 846). This fact makes it reasonable to suppose that participants do not understand the experimental questions in terms of inner feelings of action-control. Instead, when asked how much they felt they controlled the action, it is very likely that participants are assessing how much the experimental situations were similar to a situation in which they are themselves moving their arm. Even if there are sensory (perhaps in the shape of motor imagery) and affective consequences of watching another person's arm located where one's own ought to be, there is no *prima facie* reason for thinking that expressing such an assessment of similarity by assigning it a number involves a distinctive feeling of agency. If the control scale used in these experiments is thought to range from intentional action with preceding conscious decision to completely passive movement, then the preview condition should be rated higher than the non-preview condition since it mimics more elements of the ideal execution of consciously willed actions. Given the hypothesis that participants answer the introspective question by estimating the degree of the analogical similarity between a given set of trials and the prototypical intentional action, we would expect a pattern of results similar to the one obtained in the Wegner et al. study.

Wenke et al., 2010: This study investigated the influence of priming facilitating action-selection on the feeling of agency. Trials would begin with a fixation cross in the middle of a screen, followed by a subliminally presented prime in form of an arrow pointing right or left. Shortly thereafter a combined mask and target in the form of another arrow would appear. Depending on the direction of an arrow presented on a screen, participants had to press a button with the left or the right finger. After a brief interval, a coloured circle would appear on the screen. The prime would be either congruent or incongruent with the

target and hence with left or right button presses. The colours of the circles would code for congruency. Participants were told that the aim of the task was to find out “how much control they had over the colour-effect stimuli that followed their key press actions (29).” After each block consisting of 72 trials, participants’ feeling of control was measured by “a combined ranking and rating procedure”. First, the ranking procedure, participants had to sort all the presented colour effects by degree of control over the colour. In this part, every presented colour had to be assigned a rank, even if participants felt nothing or were unable to remember having seen it. Second, the rating procedure, for each colour, participants rated how much control they had on a 0 to 100 scale. The prediction was that a congruent prime should facilitate action selection and therefore enhance the feeling of agency or control. The results showed that participants tended to give colours from congruent trials a higher control rating than colours from incongruent trials. As in the Wegner et al. study, participants in this study seem to be able to introspect the strength of their feeling of agency and express it in the combined ranking and rating procedure. If that is indeed what the participants are doing, then we would have to accept that the data supports the FoA-hypothesis better than the simpler hypothesis.¹⁹

Alternative interpretation: There are a number of plausible alternative interpretations. Here I will discuss three different but possibly complementary interpretations, none of which makes any reference to a feeling of agency. According to the first interpretation, the data can be explained as a non-conscious priming phenomenon. Recall that, in the Wenke et al. study, participants were asked to rank colours with respect to the feeling of control and they had to do this for all colours, even if they did not remember the colour or any associated feeling. This is comparable to a forced choice in a visual task. If we accept that a forced choice can be biased by priming in the absence of conscious awareness of the priming stimulus, we should

¹⁹ $p(\text{data} | \text{FoA-hypothesis}) > p(\text{data} | \text{simpler hypothesis})$

accept it as likely, in the Wenke et al. study, that the primes somehow bias the ranking of colour items even in the absence of any distinctive feeling of agency. It is likely that in many cases the participants were not reporting any feeling but rather guessing by “blindly” ranking the colours by sorting them into a particular sequence. On this interpretation, the experimental manipulation facilitates action-selection that subsequently non-consciously biases the sorting of colour chips. Given the hypothesis that performance in this task can be explained by motor-control modules and non-conscious biasing of action-selection, we would expect results similar to the results found by Wenke and colleagues. The subsequent explicit rating of each colour by the participant is perhaps more difficult to explain in this way.

The second interpretation is different as it allows that participants are in fact reporting some conscious experience, only it is not a feeling of agency. The rating results could be explained as a general fluency phenomenon. Priming an action and thereby facilitating action-selection produces fluency. Maybe fluency simply feels good, so the more fluent the action selection is, the better it feels. If asked to rate colours in terms of positive affective qualities, we would predict the same results. Whether it is non-conscious biasing of the ranking of colours or biasing of the explicit rating by positive affective quality, the conclusion would be the same: there is no reason to assume the existence of a distinct feeling of agency to explain the results.

A third way to explain the reports in the Wenke et al. study would be in terms of how confident participants are that they are solving the task correctly. An explanation like this is suggested by a study by Metcalfe and Green (2007). Here participants played a game in which the task was to touch downward scrolling Xs and avoid touching Os. Participants’ control of the cursor was manipulated by turbulence of the cursor and how close the cursor had to come to the target for a hit. Performance was manipulated in other ways as well. After each trial, participants made agency judgements and judgements of performance. The authors

found that one factor consistently correlated with agency judgements was people's metacognitive assessment of their performance. The authors conclude: "these experiments revealed that one factor that contributes to people's JoA [judgements of agency] is people's assessment of their performance and that this, in turn, is closely but not invariably related to their performance itself" (Metcalf and Greene, 2007, 195). To be sure, in this study, participants were not asked to report introspectively on feelings of agency but rather asked to make agency judgements. Nevertheless, it gives us reason to suppose that metacognitive assessments of performance could be playing a role when people are asked to rate their feeling of agency. It is possible that what people report as degrees of feeling of agency is their degree of confidence in their assessment of their own behaviour. In the Wenke et al. study, ratings and rankings of colours might be an expression of participants' confidence that they controlled the colour effect. Here, participants might really be answering the question "how confident are you that you controlled the colour effect?". When answering such questions no inner feeling of control needs to be consulted. Rather, participants have to assess their willingness to bet on having made the correct assessment of their performance. This third interpretation of the results is too abstract and speculative to afford any concrete predictions. It does, however, point the way to new questions and studies. The three alternative interpretations can be combined in various ways. For example, it is possible that non-conscious priming and conscious fluency feelings are two distinct processes biasing the subject's confidence rating.

In the two discussed studies, various action relevant dimensions are being manipulated (preview vs. no-preview, self vs. no-self, action-selection-facilitation vs. no-action-selection-facilitation) and the participants' reports are treated as the dependent variable. In both studies, it is assumed that the reports are a direct measure of the feeling of agency such that the influence of various manipulations on the feeling of agency can be

studied by asking people to report their feeling of agency. I have considered a number of reasons for rejecting this assumption. There is at least one more reason for doubting that participants in the described experiments are reporting directly on their distinctive feelings of agency. This time it is a general methodological worry.

There is a methodological problem concerning memory. In both studies, reports were collected only after each block of trials. The assumption is, according to Wenke et al. (2010, 32), that participants “monitored sense of control on each trial, and accumulated – possibly imperfectly – their experience over the entire block”. On this basis, the authors assume that “numerically higher ranks and ratings indicate more perceived control” and that “participants felt more in control” (31). Is this assumption justified?

It is well known that proprioceptive and tactile signals decay very quickly (Blakemore, Wolpert and Frith, 1998). Not only are these signals difficult to attend to (Fourneret and Jeannerod, 1998), they are difficult to maintain in memory and remember even after a short time. According to the FoA-hypothesis, the feeling of agency should be comparable in these respects to proprioceptive and tactile sensations. Participants in the experiments under discussion are thus not reporting on their current experience but on the quickly decaying and fleeting experiences had over many trials, recalled after an interval of varying length. The judgements made by participants are at best a recollection, summation, and evaluation taking into account many different sources of information.

Contrast the above two studies with the demand for strict criteria in research on visual consciousness. In the visual domain, researchers often discard reports as direct evidence for or against visual experience of a stimulus if there is the slightest delay or possibility of conscious evaluation.²⁰ That is, in the visual domain, researchers are aware that in order to

²⁰ For example, after an interval, the inability to report a stimulus cannot be taken as proof of absence of experience because the subject could simply have forgotten. See, for instance, discussions about overflow of conscious experience in the Sperling-paradigm, Sperling, 1960, Block, 2011, Phillips, 2011.

gauge visual experience, there is a methodological problem of how to avoid reconstructive memory processes intervening between time of stimulus presentation and time of report. In the motor domain, this problem is exaggerated by the unfamiliarity of the tasks, methodological difficulties in experimentally controlling the object of the reports, and the praxis of letting participants report only after a completed block of trials. The fact that the participants might not remember the experiences they were having at the time of the experimental manipulation only increases the strangeness of the experimental questions and thereby the risk of participants adopting alternative interpretations of the “introspective” task instructions.

Neither of the discussed studies makes any effort to alleviate this methodological worry or differentiate and control for these various alternative interpretations of the behavioural data. If I am right in assuming that the alternative interpretations are just as likely as the FoA-interpretation, we have reason to believe that the data from the experiments are equally probable given either of our two conflicting hypotheses. To conclude, the data from these studies are not more probable given the FoA-hypothesis than given the simpler hypothesis. Without independent reasons, we should prefer the simpler hypothesis.

6. Phenomenology of Agency

It could be that there are non-experimental reasons for assuming the FoA-hypothesis. Maybe it is unnecessary to demand experimental data in support of the claim that voluntary motor actions are associated with a distinctive, “sparse intention-free” (Bayne and Pacherie, 2007, 488) feeling of agency. Maybe it is just a brute phenomenological fact. Demanding that science delivers evidence for the FoA-hypothesis would then be like demanding that pain researchers produce evidence for the claim that people experience pain.

A phenomenological argument to support the FoA-hypothesis would need a way to highlight or foreground the relevant phenomenal quality. One way to do this would be to contrast an episode with the experience in question with an otherwise similar episode where the experience is missing. In the literature, this time predominantly philosophical, we find two contrastive, phenomenological strategies for arguing in favour of a distinctive phenomenology.

The first contrastive strategy focus on everyday experiences of active and passive movements. One way to bring out or foreground the relevant quality could be to contrast two cases that differ only with respect to the quality in question. Contrast, for example, the case of intentionally moving your arm from A to B with the case of your arm being pushed or pulled from A to B. Imagine that the kinematics of the active and passive movements are the same. It seems that in both cases, you have an experience of your arm's movement from A to B, but only in the first case would you have an experience of initiating and controlling moving your arm from A to B. It seems that the relevant difference between the two cases is the feeling of agency, the specific sense of initiating and controlling the action (for such contrastive scenarios, see Bayne, 2011, Gallagher, 2000).

Everyone can agree that one difference between the two cases is that in the first case the agent is moving her arm intentionally from A to B, whereas in the second the agent is not. Everyone can equally agree that it is true to say that in both cases the agent's arm is moving from A to B. Given normal constraints on attribution of intentional agency, most people should accept that in the first case the agent is aware of moving her arm from A to B intentionally. It is, however, a further thing to claim that this difference should be explained in terms of a distinctive, intention-free feeling of agency. There is no reason to assume that any such thing is being foregrounded in this contrast. Maybe what is being foregrounded is the difference between things going as expected and things not going as expected – surprise

with a negative valence. In fact, many qualities could change across the two cases. When my arm is being pushed, I have a tactile experience of being pushed. And when I am moving voluntarily, I know what I am doing and trying to do, and this might involve a variety of experiences (of a proprioceptive, visual, haptic, etc., kind). It is a moot question how to spell out the agent's awareness of what she is doing, but it is not generally assumed in the philosophical literature that it can be explained in terms of a distinctive intention-free feeling of agency tied to comparator-based motor control processes. It is an equally moot question, which sensory experiences might be involved, how they are involved, or whether they really do remain unaltered across the contrast cases. This contrast between moving voluntarily from A to B and being moved from A to B is consequently insufficient to foreground a simple and distinctive qualitative difference.

Maybe I am considering the wrong type of contrast case. To be sure, the argument here works by elimination. If successful, it only shows that some contrasts fail in foregrounding a feeling of agency. The one I have just considered is, however, the one that is being considered in the literature on the subject. The onus is therefore on the proponent of the FoA-hypothesis to produce a contrast scenario convincingly highlighting only the distinctive feeling of agency.²¹

This brings me to the second contrastive strategy. One might think that the problem here is that two different categories are being compared: intentional actions and involuntary movements. To bring out the qualitative feeling of agency we would need to contrast a case of action with the relevant feeling and a case of action where the feeling is for some reason significantly altered. Authors contrasting the normal case of intentional action with various

²¹ One reviewer suggested that better cases to discuss would be Wegner's cases of automatisms (Wegner, 2002, Ch. 1 & 4). These are cases where an apparent voluntary movement is performed without any explicit intention or even with explicit disavowal of intention. These are, however, also cases performed without any feeling of agency, if any such exist. Since intentions and feeling of agency disappear together in these cases, they cannot serve as evidence for an intention-free feeling of agency. Furthermore, most of the descriptions are anecdotal. The subjective reports vastly underdetermine our choice of phenomenological description.

pathologies of agency, e.g. patients suffering delusions of alien control, have frequently exploited this version of the contrast strategy. One version of this strategy would go something like this. A patient experiencing alien control might decide to raise her arm, might be aware of her intention to be raising her arm, and might be aware of her arm's rising. Nevertheless, she might still deny that she is the one raising her arm and attribute the control of the arm's rising to some external agent. This strange misattribution of control is then explained by the patient's lacking something normal agents have: the feeling of agency (see, for example, Bayne and Pacherie, 2007, Bayne, 2011).

This explanatory strategy is confronted with an immediate problem. From the fact that agency is somehow experienced as lost or challenged in these syndromes it does not follow that we have a full positive experience of agency in the normal case. The way in which we describe these cases can be deceptive. If delusions of alien control and other such disturbances are described as an experience of loss of control or diminished control, then you might think that in the normal case we would have simply an experience of control (it would be a full sense of control, nothing would be lost). This is a consequence of describing the negative case as one where something is lost that is there in full figure in the positive case. The situation might not be like this.

What might the situation be like, then? How a specific type of pathological action is best described is of course an important and complex issue. In a full coverage of this issue, we would need to look systematically at reports made by patients. My conjecture is that the various types of pathological cases are not best described as a loss or alteration of a *normal intention-free feeling of agency*. Take the case of schizophrenic patients' delusions of alien control. There are a number of reasons for thinking that a contrast between the normal experience of action and disturbed experiences involved in the delusion would yield no clear result. The very idea of such a contrast builds on a number of controversial assumptions.

Here I will briefly highlight some of the most controversial or problematic assumptions, not to demonstrate their falsity, but merely to make it manifest that there is no simple or direct way to compare the normal and delusory experiences of action and infer the presence or absence of a single phenomenal aspect.

First, even if we accept a cognitive neuropsychiatric approach (e.g. Frith & Done, 1988, Coltheart, 2007, Coltheart, Langdon & McKay, 2011), the contrast between normal and pathological action need not be informative with respect to phenomenal feelings of agency.²² In order for the contrast to be informative in this respect, it must be assumed that the schizophrenic agent's delusional belief and judgements are grounded on conscious delusional experience. It must be assumed that the delusion of alien control involve a low-level phenomenal feeling involved in action. This assumption is by no means unanimously accepted. In a recent commentary, Wong (2012, 50) expresses doubts concerning this epistemological picture. Reflecting on the puzzling judgements made in cases of pathologies of agency, he observes: "when we consider the puzzling pattern of deficits that we encounter in various pathological conditions, such as schizophrenia, it becomes very unclear that the judgement of agency must be read off from the feeling of agency, but cannot be directly generated by underlying mechanisms." And continues: "There is no reason why a delusional agency belief or judgement in a schizophrenic cannot be directly generated by predictive neural processes without these processes first generating an anomalous agency experience."²³

Second, even if we concede this point (delusions of alien control are rational responses to or endorsements of specific experiences had by the schizophrenic agent), it still does not follow that the contrast between the schizophrenic agent's experience and the normal agent's awareness of action will tell us anything about the normal agent's feeling of

²² The neurocognitive approach to psychopathology is of course not universally accepted. For another influential but fundamentally conflicting approach, see Bovet and Parnas, 1993, Parnas, 1999, and Parnas et al., 2005.

²³ A similar view is endorsed by Coltheart et al. 2010, 264 (in an explanation of Capgras delusion): "the first delusion-relevant event of which the patient is aware is the belief".

agency. It might be the case that the delusional agent's experience should be explained by a deficit of dedicated motor control mechanisms; but it might equally be the case that only in the case where this mechanism is malfunctioning does it give rise to a conscious signal. When the comparator-based motor control mechanisms are working flawlessly, they do not produce any conscious output. In fact, one of their functions might be dampen down or cancel out sensations involved in moving one's body. So, one suggestion might be that when these predictive mechanisms or comparators are malfunctioning, then no cancelling out of sensations will take place, and self-initiated movements will begin to feel like "unpredictable passive movements" (Frith & Johnstone, 2003, 138). The delusional agent might explain these experiences by adopting the delusional belief that an alien force is controlling the movements.²⁴ In order for the contrast between normal experience of agency and delusion of alien control to yield any information about the normal feeling of agency, it must be assumed that the disturbing experience of agency involved in the delusion is an alteration of a normal feeling of agency. The upshot of this discussion is, however, that neither the explanation of voluntary action by inverse and forward models and comparator-mechanisms nor the cognitive neuropsychiatric approach support this assumption.²⁵

To conclude, neither of the two strategies for contrasting experiences succeeds in foregrounding a simple phenomenal quality. There are no simple phenomenological facts concerning the feeling of agency to be gained from these contrastive exercises. Phenomenological reflection does not render the FoA-hypothesis more probable.

²⁴ Similar view is expressed by Coltheart, 2005, 155: "Only when a prediction fails does consciousness get involved; the unconscious system makes some kind of report to consciousness to instigate some intelligent conscious problem-solving behavior that will discover what's wrong."

²⁵ The argument works by exclusion. Maybe there are contrast cases I have not considered that are genuinely informative. Some authors also use the case of Anarchic Hand Syndrome (e.g. Bayne, 2011). This type of case can easily be dismissed. Since both intention and (alleged) feeling of agency are missing in AHS, they are not informative with respect to an intention-free feeling of agency. See, for example, Cheyne, Carriere, and Smilek, 2009, 481: "[...] the intentions do not seem to be their concurrent intentions and hence agency is transferred from self to hand. Such patients have experienced not merely a loss, but an active alienation, of agency for the anarchic actions". See also, Kritikos, Breen, and Mattingly, 2005, and Della Sala, 2005.

7. Concluding Remarks

The FoA-hypothesis is the hypothesis that the performance of voluntary movements is associated with a low-level, comparator-based, distinctive and intention-free feeling of causing one's own movements. Distinctiveness and intention-freedom go together: the comparator-based feeling of agency is distinctive, which implies that it cannot be reduced to an experience of consciously intending to do something. According to the FoA-hypothesis, the distinctive and intention-free feeling of agency is a defeasible subjective mark of voluntary agency and is therefore apt to serve as the epistemic ground for agency judgements. That is, in the normal case, it is rational to judge that you are causing some movement or event only if you experience a feeling of agency for that movement or event.

In this paper, I have treated the FoA-hypothesis as a scientific hypothesis and asked what type of evidence we have to support it. In order to answer this question, I have contrasted the FoA-hypothesis with a simpler hypothesis that is identical to the FoA-hypothesis except for the fact that according to the simpler hypothesis the comparator-based motor control system does not output a distinctive feeling of agency. I then looked at two different types of experimental paradigms often supposed to study the feeling of agency (viz. action-recognition and "introspective" paradigms), and asked whether data from these paradigms are more likely given the FoA-hypothesis than given the simpler hypothesis. It was found that data from these experiments are equally likely according to the FoA-hypothesis and the simpler hypothesis. Without independent reasons for assuming the existence of a low-level, comparator-based, distinctive and intention-free feeling of causing one's own movements, we would have no reason to accept the FoA-hypothesis.

Could it be that I have misrepresented the FoA-hypothesis? Maybe the feeling of agency is not distinctive in the sense of being a phenomenally specific feeling of causing

one's own movement but rather a distinctive kind of pattern in the flow of perceptual experiences. The latter view is defended by Prinz (2007). According to Prinz, the forward model of the comparator-based control system predicts sensory consequences of the movement, and when the brain is processing the reafferent feedback from performing the movement, the predicted feedback is cancelled out or attenuated. When sensory feedback is unpredicted or does not match the prediction, it stands out as something the agent needs to pay attention to.²⁶

Nothing I have said contradicts the claim that comparator-based control systems generate a distinctive pattern in the flow of perceptual experience. This being said, it is worth noticing that on this account of the comparator mechanisms' contribution to perceptual phenomenology, the experience generated by comparator mechanisms no longer serves two functions crucial to the FoA-hypothesis. First, the comparator-based perceptual flow experience could not serve as a mark of voluntary agency. Even if the comparator mechanisms of motor control explains this pattern of flow of anticipations and feedback only with respect to the movements, we find the same structure of ongoing flow of anticipation and sensory feedback in other domains as well, such as speech processing and perceptual learning. We have the structure of anticipation and sensory feedback when observing other people's actions, listening to music, and going for the nth ride in a roller coaster. Second, as a consequence, there would no longer be any rationale for claiming that the feeling of agency serves as the epistemic ground for agency judgements. The feeling of agency would not provide me with a reason for thinking that I am the cause of some event. This should suffice to show that proponents of FoA-hypothesis do not have this "perceptual flow" interpretation in mind.

²⁶ See Pacherie, 2001, 174: "our ordinary experience of agency may simply be a rather diffuse sense of a coherent ongoing flow of anticipations and sensory feedback." Also Hohwy and Frith, 2004, 185.

This conclusion does not exclude perceptual anticipations and flow from playing distinctive roles in the agents' knowledge of their own actions and their agency judgements according to the simple hypothesis. Here, one suggestion could be that when sensory anticipations and experiences are related in the right way to the agent's intention, sensory evidence could be having an important epistemic role in explaining and justifying agency judgements.²⁷ Whether the comparator model of motor control in the end turns out to be the best explanation of motor related sensory anticipations and flow-experiences is not crucial to this epistemic account of agents' knowledge and agency judgements.

If we reject the FoA-hypothesis, what is then the alternative? According to a dominant view, there are two possible and competing models of agentic experience. The first model is the FoA-hypothesis. The second model is an inferential model according to which agentic experience depends on holistic, domain-general mechanisms involved in "narrative" self-understanding (Stephen and Grahams, 2000, Wegner, 2002). According to the most prominent version of this model, the experience of agency is the product of a post hoc, reconstructive inference from prior thoughts and post-action sensory events. As Haggard and Tsakiris put it: "According to [the inference model], the sense of voluntary control is effectively a reconstructive illusion that one's intention has caused an external event, analogous to Hume's view of causation as an illusory inference from the constant conjunction of cause and effect" (2009, 243). By arguing against the FoA-hypothesis am I then arguing for the inferential reconstructive model?

The answer is no. The choice between the FoA-hypothesis and the inferential-reconstructive model is not exclusive. There are other options. Recall that the simple hypothesis is exactly like the FoA-hypothesis except for the feeling-component. The simpler hypothesis thus involves conscious intentions, conscious anticipations, awareness of what one

²⁷ See Grünbaum, 2011, for an argument in support of the idea that perception often has a distinctive, epistemic function in agents' knowledge of their own actions.

is trying to do, and various bodily and perceptual experiences. We are not forced to claim that the agent's awareness and agentic experiences are inferred only after performing the action. The agent's experience could be part of her conscious intention to be doing the action or could be a consequence of having such an intention, and as such this experience might predate or be concurrent with the performance. According to a prominent philosophical model of agents' knowledge of their own action, an agent is normally successful in her skill-based actions. When she intends to perform a skill-based action, she is normally in a position to know what she will be doing or is doing (depending on type of intention, for discussion, see Falvey, 2000, Grünbaum, 2011). Such an account is in line with some researchers' emphasis on premotor determinants of the agentic experience. According to this emphasis, the agentic experience is to a large extent dependent on premotor components of action, such intention, planning, and action-selection (Vinding et al., 2013, Jensen et al., 2013, Chambon and Haggard, 2013). Denying that there is a distinctive, intention-free feeling of agency generated by low-level comparator-based control processes is therefore neither to deny that there are distinctive mental states involved in acting intentionally, nor to deny that comparator-based control mechanisms are important to control of voluntary movements.

Acknowledgements

I am grateful to Adrian Alsmith, Glenn Carruthers, Mark Schram Christensen, Ophelia Deroy, Frederique de Vignemont, Mads Jensen, John Michael, Krisztina Orban, Barry Smith, Mikkel C. Vinding, Martin Voss, Hong Yu Wong, and a number of anonymous reviewers for their helpful comments to earlier versions of the manuscript. The material has been presented at workshops and conferences in Aarhus, Antwerp, Copenhagen, Granada, London, and Tübingen, and I have received many helpful comments and suggestions. Research for this

paper was funded by a grant from the Danish Research Council, FKK, project “Phenomenal Consciousness and Cognitive Motor Control”.

References

- Bar-On, D. (2004). *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Oxford University Press.
- Bayne, T. (2011). The sense of agency. In F. Macpherson (ed.), *The Senses*. Oxford: Oxford University Press.
- Bayne, T. and Levy, N. (2006). The Feeling of Doing: Deconstructing the Phenomenology of Agency. In N. Sebanz and W. Prinz (eds.), *Disorders of Volition*. Cambridge, MA: MIT Press.
- Bayne, T., & Pacherie, E. (2007). Narrators and comparators: the architecture of agentic self-awareness. *Synthese*, 159(3), 475-491.
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, 1(7), 635-640.
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15(12), 567-575.
- Bovet, P., & Parnas, J. (1993). Schizophrenic delusions: a phenomenological approach. *Schizophrenia Bulletin*, 19(3), 579.
- Butterfill, S. A. & Sinigaglia, C. (2014), Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*, 88: 119–145.
- Carruthers, G. (2012). The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and cognition*, 21(1), 30-45.

- Chambon V. & Haggard P. (2013). Premotor or ideomotor: How does the experience of action come about? In Prinz, W., Beisert, M. and Herwig, A. (eds.), *Action Science: Foundations of an Emerging Discipline*. Cambridge, MA: MIT Press.
- Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex*, 23(5), 1031-1037.
- Cheyne, J. A., Carriere, J. S., & Smilek, D. (2009). Absent minds and absent agents: Attention-lapse induced alienation of agency. *Consciousness and cognition*, 18(2), 481-493.
- Coltheart, M. (2005). Conscious experience and delusional belief. *Philosophy, Psychiatry, & Psychology*, 12(2), 153-157.
- Coltheart, M. (2007). Cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology*, 60(8), 1041-1062.
- Coltheart, M., Menzies, P., & Sutton, J. (2010). Abductive inference and delusional belief. *Cognitive neuropsychiatry*, 15(1-3), 261-287.
- Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual review of psychology*, 62, 271-298.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., and Jeannerod, M. (1997). Looking for the agent. An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71–86.
- Daprati, E., Wriessnegger, S., and Lacquaniti, F. (2007). Kinematic cues and recognition of self-generated actions. *Experimental Brain Research*, 177(1), 31-44.
- Della Sala, S. (2005). The anarchic hand. *The Psychologist*, 18(10), 606-609.
- Falvey, K. (2000). Knowledge in Intention. *Philosophical Studies*, 99, 21-44.

- Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *Neuroimage*, 18(2), 324-333.
- Farrer, C., Frey, S.H., Van Horn, J.D., Tunik, E., Turk, D., Inati, S., and Grafton, S.T. (2008). The angular gyrus computes action awareness representations. *Cerebral Cortex*, 18(2), 254-261.
- Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36(11), 1133-1140.
- Frith, C. D., & Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *The British Journal of Psychiatry*, 153(4), 437-443.
- Frith, C. & Johnstone, E. (2003). *Schizophrenia*. Oxford: Oxford University Press.
- Friston, K. (2011). What is optimal about motor control?. *Neuron*, 72(3), 488-498.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14-21.
- Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, 30(1), 15-31.
- Grünbaum, T. (2011). Perception and non-inferential knowledge of action. *Philosophical Explorations*, 14(2), 153-167.
- Haggard, P., & Johnson, H. (2003). Experiences of voluntary action. *Journal of Consciousness Studies*, 10(9-10), 72-84.
- Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, 22(10), R390-R392.
- Haggard, P., & Tsakiris, M. (2009). The Experience of Agency Feelings, Judgments, and Responsibility. *Current Directions in Psychological Science*, 18(4), 242-246.
- Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche*, 13(1), 1-20.

Hohwy, J. and Frith, C. (2004). Studies of the neural correlates of consciousness can do better, but are on the right track. *Journal of Consciousness Studies*, 11(1): 45–51.

Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford: Blackwell Publishing.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford: Oxford University Press.

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a reappraisal. *Experimental Brain Research*, 192(3), 527-532.

Jensen, M., Vagnoni, E., Overgaard, M., & Haggard, P. (2013). Experience of action depends on intention, not body movement: An experiment on memory for mens rea. *Neuropsychologia*.

Kritikos, A., Breen, N., & Mattingly, J. B. (2005). Anarchic hand syndrome: Bimanual coordination and sensitivity to irrelevant information in unimanual reaches. *Cognitive Brain Research*, 24, 634–647.

Kühn, S., Brass, M., & Haggard, P. (2013). Feeling in control: Neural correlates of experience of agency. *Cortex*, 49(7), 1935–1942.

Machery, E. (2013). In Defence of Reverse Inference. *The British Journal for the Philosophy of Science*.

Metcalf, J., and Greene, M. J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General*, 136(2), 184-199.

Moore, J., and Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and cognition*, 17(1), 136-144.

Pacherie, E. (2001). Agency lost and found: a commentary on Spence. *Philosophy, Psychiatry, & Psychology*, 8(2), 173-176.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107(1), 179-217.

- Parnas, J. (1999). From predisposition to psychosis: progression of symptoms in schizophrenia. *Acta Psychiatrica Scandinavica*, 99(s395), 20-29.
- Parnas, J., Handest, P., Jansson, L., & Sæbye, D. (2005). Anomalous subjective experience among first-admitted schizophrenia spectrum patients: empirical investigation. *Psychopathology*, 38(5), 259-267.
- Phillips, I. B. (2011). Perception and Iconic Memory: What Sperling Doesn't Show. *Mind & Language*, 26(4), 381-411.
- Prinz, J. (2007). All consciousness is perceptual. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 335–357). Oxford, England: Blackwell.
- Ritterband-Rosenbaum, A., Christensen, M. S., Kliim-Due, M., Petersen, L. Z., Rasmussen, B., & Nielsen, J. B. (2011). Altered sense of Agency in children with spastic cerebral palsy. *BMC neurology*, 11(1), 150.
- Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition*, 110(1), 74-83.
- Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, 94(3), 241-255.
- Sober, E. (1999). Testability. In *Proceedings and Addresses of the American Philosophical Association* (Vol. 73, No. 2, pp. 47-76). American Philosophical Association.
- Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge University Press.
- Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain*, 120(11), 1997-2011.

- Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, 74(11), 1-29.
- Stephens, G. L., & Graham, G. (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. The MIT press.
- Stirling, J. D., Hellewell, J. S., & Quraishi, N. (1998). Self-monitoring dysfunction and the schizophrenic symptoms of alien control. *Psychological medicine*, 28(03), 675-683.
- Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and cognition*, 17(1), 219-239.
- Tsakiris, M., & Haggard, P. (2005). Experimenting with the acting self. *Cognitive Neuropsychology*, 22(3-4), 387-407.
- Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, 96(3), 215-231.
- Van Den Bos, E., & Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition*, 85(2), 177-187.
- Vinding, M. C., Pedersen, M. N., & Overgaard, M. (2013). Unravelling intention: Distal intentions increase the subjective sense of agency. *Consciousness and cognition*, 22(3), 810-815.
- Wegner, D. M. (2002). *The illusion of conscious will*. MIT press.
- Wegner, D. M. (2004). Précis of The illusion of conscious will. *Behavioral and Brain Sciences*, 27(5), 649-659.
- Wegner, D. and Wheatley, T. (1999). Apparent Mental Causation: Sources of the Experience of Will. *American Psychologist*, 54(7), 480-492.
- Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of personality and social psychology*, 838-848.

Final draft. To appear in *Synthese*. **Please** Cite Published Version.

Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, 115(1), 26.

Wong, H.Y. (2012). A Measure of My Agency? *Consciousness and Cognition*, 21, 48-51.