

SENSE OF AGENCY DOES NOT SUPPORT ANTI-REDUCTIONIST THEORIES OF ACTION*

A SENSACÃO DE AGÊNCIA NÃO APOIA TEORIAS DE AÇÃO ANTIRREDUACIONISTAS

Beatriz Sorrentino Marques

<http://orcid.org/0000-0001-7192-0777>

bsorrentinom@gmail.com

Universidade Federal de Mato Grosso, Brasil

ABSTRACT *Anti-reductionist Theories of Action are, in general, accounts of action according to which the role the agent plays in the production of her action cannot be reduced to her mental events or states. Proponents of these theories present claims according to which Anti-reductionist Theories of Action better capture the sense of agency that agents experience when performing intentional actions than the competing Causal Theory of Action. I will investigate the arguments supporting these claims and argue that they do not survive scrutiny, especially when taking into consideration the neuro-cognitive information available about the sense of agency. Anti-reductionist Theories of Action seem incompatible with the empirical account of sense of agency, as show experiments focusing on action selection, the comparator model, and perception of the relation between action and its outcome. Alternatively, the*

* Article submitted on 19/02/2020. Accepted on 14/10/2020.

Causal Theory of Action is more able to accommodate and explain the sense of agency.

Keywords *sense of agency, comparator model, action selection, control, reductionism.*

RESUMO *Teorias da Ação Anti-reducionistas são, de modo geral, explicações para ações de acordo com as quais o papel que o agente desempenha na produção de sua ação não pode ser reduzido aos seus eventos e estados mentais. Proponentes dessas teorias apresentam afirmações de acordo com as quais as Teorias da Ação Anti-reducionistas capturam a sensação de agência que agentes experienciam quando empreendem uma ação intencional melhor do que a concorrente Teoria Causal da Ação. Investigaremos os argumentos que apoiam essas afirmações e defenderemos que eles não sobrevivem ao escrutínio, especialmente quando levamos em consideração a informação neuro-cognitiva disponível a respeito da sensação de agência. Teorias da Ação Anti-reducionistas parecem incompatíveis com a explicação empírica da sensação de agência, como mostram experimentos focados na seleção de ações, no modelo comparador e na percepção da relação entre ação e seu resultado. Já a Teoria Causal da Ação é mais capaz de acomodar e explicar a sensação de agência.*

Palavras-chave *sensação de agência, modelo comparador, seleção de ação, controle, reducionismo.*

1. Introduction

The *sense of agency* the agent experiences when she performs an intentional action has become a point of interest for both neuroscience and philosophy. Proponents of Anti-reductionist Theories of Action (Ginet, 1997; Lowe, 2008; Nida-Rümelin, 2007; O'Connor, 1995, 2009; Pereboom, 2015) claim that sense of agency supports their theories. Here, I will investigate this kind of claim. In section two, I characterize sense of agency. I discuss the scientific account of sense of agency in section three. In section four, I argue that the scientific account does not support Anti-reductionist Theories of Action (ATA) and that it favors the Causal Theory of Action (CTA). Finally, in section five, I respond to possible objections.

2. Sense of agency

There are three ways to investigate the sense of agency: conceptually, phenomenologically, and neurocognitively. The conceptual analysis focuses on the conceptual elements encompassed by the sense of agency, but I will not focus on it here. I will focus on the phenomenological aspect and whether ATA's interpretation of it is supported by a prominent neurocognitive theory and its empirical data.

On the phenomenological analysis, the sense of agency is distinguished from the sense of ownership. In addition, their pre-reflective aspects are distinguished from their reflective aspects. I will start by elucidating these distinctions. The sense of agency is a sense or feeling that accompanies acting.¹ It can be defined as the experience that I caused the action (Gallagher, 2012). The contrast with the sense of ownership of movement, which is the experience of one's body moving, voluntarily or not, helps making the idea of sense of agency clearer. I will focus on the sense of agency.

The reflective aspect of experience involves introspective reflection, which is reflecting on one's own conscious experience, plans, or evaluations. It requires to metacognitively engage with one's mental states, cognitive processes, and conscious experience. On the other hand, situated reflection does not require metacognition or conceptualization; it is "embedded with the ongoing context of action (Gallagher, 2012, p. 17), i.e., it is pre-reflective. As it may seem intuitive, the sense of agency is pre-reflective; it is a recessive form of consciousness.

There is no doubt, nevertheless, that reflective aspects may contribute to the sense of agency (Synofzik, Vosgerau, and Newen, 2008; Gallagher, 2012). According to Pacherie, stages of intention cascade into the next in the production of action, future-directed intentions (F-intentions) cascade into present-directed intentions (P-intentions), which cascade into motor intentions (M-intentions) (Pacherie, 2008). As Gallagher draws the phenomenological distinction, F and P-intentions involve conscious mental states and may contribute to the reflective aspect of the sense of agency. However, this is not a necessary aspect for the agent to have the sense of agency; only M-intentions necessarily are associated with the pre-reflective sense of agency, for M-intentions do not involve conscious mental states; they are sensorimotor representations.

1 The present discussion takes only overt intentional action into consideration, not mental action, because it is unclear whether the motor processes involved in explaining the sense of agency can apply to mental action. It is unclear even if agents have the sense of agency when they perform mental actions.

The distinction between pre-reflective and reflective is in line with Synofzik, Vosgerau, and Newen (2008). They convincingly argue that sense of agency (they call it the feeling of agency) is the non-conceptual phenomenology of agency, originated by perceptual and motor control processes. On the other hand, interpretative judgments about agency are best-called judgments of agency and allow the agent to explicitly attribute the action to herself.

The present discussion will concentrate on the pre-reflective aspect of the sense of agency, the experience of oneself as an agent. It will not be concerned with the agent's capacity to reflectively attribute the action to herself as the cause of the movement (to which F and P-intentions contribute). I will accept that the explanation for the sense of agency has both motor and intentional elements. According to the motor explanation, "[sense of agency] is generated at the level of primary sensory-motor processes that involve efferent and reafferent processes such as motor (efferent) commands and proprioception and is manifested at the level of first-order phenomenal consciousness" (Gallagher, 2012, p. 18). Additionally, the sense of agency encompasses actions' teleology, for it is also the experience that we control events in the world, which is explained by the perceptual monitoring of whether the intended goal was achieved.

The issue of stating what the sense of agency is involves spelling out the conceptual taxonomy of elements that compose the phenomenology of acting (reference withdrawn). I will not take up this task here; it is enough to distinguish the two aspects that contribute to the phenomenology of acting: the pre-reflective and the reflective aspect. Considering that it is difficult to precisely define what the sense of agency is (Gallagher, 2012), I accept that it is the experience that the action is self-caused (Synofzik, Vosgerau, and Newen, 2008; Gallagher, 2012) and that it accompanies overt intentional action.

There is a growing empirical literature on the sense of agency, associated with the phenomenological aspect. In the next section, I will focus on some data from cognitive-neuroscience that I consider relevant to explain the sense of agency and how it is related to the production of human action. Data are discussed presupposing a theory about the sense of agency—similarly to the one accepted by Pacherie (2008) and Gallagher (2012)—, what Haggard and Clark (2003) call the constructive view. The constructive view claims that the sense of agency is constructed as a byproduct of the preparation and production of action, "conscious experience of action relies on appropriate temporal and predictive relations between preparation, movement, and effect, and does not arise from a general context or loose conjunction of these events" (Haggard and

Clark, 2003, p. 705).² My purpose in considering empirical data is to illuminate whether ATA captures this phenomenon better than a reductionist theory, CTA.

3. The scientific account of sense of agency

Although its role is still under debate (Mylopoulos, 2015; Mylopoulos and Shepherd, forthcoming), the comparator model³ aims to capture how the neuro-cognitive mechanisms that produce the sense of agency work. Roughly, it ensures that the outcome of action matches an initial motor command by monitoring and correcting movement. The idea is similar to computational models of motor control. The inverse model specifies a motor command (movement intention), arguably, the one necessary to satisfy the agent's intention by producing a goal-directed movement (Haggard, 2017, p. 202). It also generates a copy of the motor command, called an efference copy.⁴ The forward model, on the other hand, uses the efference copy to compute the sensory consequences of the motor command, i.e., the bodily movements and the outcome of the action in the world. "Sensory information about the body and the environment is then compared with the sensory feedback that would be predicted given the motor command" (Haggard, 2017, p. 201). Haggard explains that when the motor command to flick the light switch is sent, the prediction is that the lights will come on. "If the arm does not move in the appropriate way, the motor control system must update or alter the motor command to achieve the goal of switching the lights on" (2017, p. 201).

Considering that it is a control mechanism, the system may generate a prediction error⁵, or not—which is only generated when predictions do not match the feedback signals. Haggard states that "[i]nterestingly, according to the comparator model, the sense of agency is caused by the lack of any

2 In opposition to the reconstructive view—Wegner (2002) might be said to hold this view—which claims that sense of agency is an illusion based on retrospective inferences the agent makes about her action.

3 Feinberg (1978) applies to the phenomenon of *thought intrusion*, in schizophrenia, the association of motor commands with control mechanisms involving internal feedback, as well as the distinction between events that happen in the world from those that are self-produced. He proposed that thought production might be a motor process dependent on internal feedback as well, so problems with the control mechanism might result in problem in thinking. Frith (2012) acknowledges that the comparator model was first proposed to address this issue, but questions whether thought and action might be analogous. For instance, it might be hard to predict the consequences of thought, as motor movements can be predicted, because it is not clear what the consequences of thought are. Hence, in the case of schizophrenia, the comparator model is best applied to explain the illusion of control.

4 An efference copy is a copy of the outgoing motor command (going to the muscles) used by a forward model of a control mechanism to anticipate and cancel, if necessary, the sensory effects of a movement (Wolpert, 1997).

5 The difference between the outcome of the action predicted by the efference copy and the actual outcome of the action (Haggard, 2017).

prediction error, implying absence of any signal at the output of the neural comparator” (2017, p. 201). The prediction error can attenuate the sense of agency because it indicates that action does not go as expected. On the other hand, lack of prediction error allows for the rise of the sense of agency, through the contribution of other mechanisms that will be discussed below—perceived regularity between the action and its outcome can even override the prediction error signal and produce a sense of agency. Furthermore, when there is a prediction error, the comparison can be used to adjust the motor command in order to make the outcome of the action match the intention. Nevertheless, lack of prediction error is not the comparator model’s only contribution to the sense of agency; the role that M-intentions play in the comparator model and the sense of agency must not be forgotten, for intentions bind the action with its outcome.

As experiments involving *intentional binding*⁶ suggest, the comparator model’s contribution to the sense of agency seems to involve a relation to the agent’s intention (cognitive preparation for action). Haggard and Clark (2003) showed that when the agent’s intention matches her movement but is produced by Transcranial Magnetic Stimulation (TMS), she does not have the sense of agency. Therefore, a movement merely corresponding to the intention does not inferentially engender a sense of agency; the intention must play a role in producing movement in order to engender that sense of agency. According to Pacherie (2008), intentions cascade from P-intentions to M-intentions to produce the action; on the other hand, perception monitors whether the intended goal was achieved (Wen and Haggard, 2020). Outcome control is a relevant element, as well, to the sense of agency (Beck, Di Costa, Haggard, 2017).

In a recent experiment, Wen and Haggard (2020) showed that perception of the relation between an action and its results elicits the sense of agency. They tested this hypothesis by creating an experiment in which subjects perceived the result of their action in relation to the action, even in a situation in which there was a prediction error, i.e., situations in which the comparator model would not give rise to the sense of agency.

In one of Wen and Haggard’s (2020) experiments, roughly, the task was to point out which of three dots on the screen the subject could control, four interference conditions were applied to the task. One of the disturbances was

6 Intentional binding is observed in intentional actions. The agent tends to perceive her action shifted forward in time towards the outcome of the action, and the outcome backwards in time towards the action (Haggard and Clark, 2003). This effect does not occur in cases of unintentional movement, and it is considered an implicit indicator of the sense of agency.

mixing either 60% or 90% of another participant's prerecorded mouse movement with the subject's movement. The other was a 30° or 90° clockwise angular bias in the transformation of mouse movement onto the dot on the screen. The experimenters had already observed in a motor control experiment that the 60% and the 90° interferences ensued the *same amount of considerably reduced motor control* to the agent (the 30° interference resulted in good and the 90% in extremely poor control). The most remarkable result concerns those interference conditions that resulted in reduced control. In the 60% condition, subjects could not keep track of the relation between movement and visual outcome, but in the 90° interference condition, they could, even though they had similar motor control in the two conditions. Wen and Haggard suggest that the subjects' sense of agency was preserved in the 90° condition. The preservation was due to the computation of regularity between movement and outcome on the screen, which would explain how subjects could tell which dot they were moving in the 90° condition.

All four interference conditions involved subjects moving the mouse, but a sense of agency did not accompany all actions. This point is apparent in the two conditions that allowed the same amount of motor control. The errors in control would produce a large prediction error, but in the case in which subjects perceived the regularity between their action and its outcome, their actions still seemed accompanied by a sense of agency, showing that perception also contributes to the sense of agency.

Additionally, Haggard (2017) suggests that a higher probability that the outcome of the action will match the agent's intended outcome increases the sense of agency. Moreover, the fluency⁷ in selecting which action the agent will perform raises the probability that the agent will feel in control of which outcomes she produces. One piece of evidence of the contribution of action selection to the sense of agency is that when different action courses are available to choose from, for instance, seven buttons to push instead of one in an experimental setting, the availability of alternatives significantly enhances *intentional binding* (Barlas and Obhi, 2013).

Action selection fluency seems to directly contribute to the sense of agency, independently of monitoring the action outcome, or of whether the action reaches the intended outcome (Haggard, 2017). One suggestion is that the metacognitive signals generated in action selection contribute to the sense

7 Fluency refers to how easily the action selection mechanism selects which action will be performed out of the agent's repertoire to best reach the relevant goal. It is associated to certainty about which action to perform (Haggard, 2017).

of agency as experiments with subliminal priming show. Wenke, Fleming and Haggard (2010) developed an experiment in which subjects had to choose between a right and a left key press. The experimenters used subliminal priming—a left or right-pointing arrow target—to manipulate the fluency of action selection; thus, action selection could be either consistent with the prime or inconsistent. The priming influenced action selection fluency. Finally, a color patch appeared on screen as the result of the keypress, but the relation between the keypress and which color appeared was not predictable based on action, prime, or target. The colors were linked to the compatibility between the prime (of which subjects were not consciously aware) and the action performed.

Outcome monitoring, therefore, cannot account for the sense of agency in this case, because the outcome did not depend solely on the action performed; it depended on prime compatibility with action. Thus, it was unpredictable, based on which action the subject performed. Participants reported a stronger sense of agency when the action selection was compatible with the prime. The experiment showed a dissociation between processes of action selection and outcome matching, leading the experimenters to attribute the results to the facilitation of action selection, not to the predictability of action results.

In order to rule out the possibility that response time (fast response) would account for a more robust sense of agency in Wenke, Fleming and Haggard's (2010) experiment, Chambon and Haggard (2012) increased the temporal distance between the mask⁸ and the task-relevant target (a right or a left arrow that subjects could see) in a similar experiment. Considering that greater fluidity in action selection would also mean faster response time, the experimental protocol dissociates action selection fluency from response time because although priming usually increases response time, it slows response time down in cases in which mask and target are distant in time⁹ (see Chambon and Haggard, 2012, p. 442). Subjects reported a strong sense of control¹⁰ in trials in which the prime was compatible with the target arrow presented, even though their responses were slower in those trials. The experimenters concluded that this is evidence that action selection¹¹—not performance monitoring—contributes to a sense of control.

8 The right or left arrow presented below conscious threshold.

9 Known as the negative compatibility effect (NCE).

10 Subjects had to answer how much *control* they had over the color patch presented on screen at the end of each trial, but Chambon and Haggard (2012) draw conclusions about the sense of agency. Although I disagree with the assumption that the sense of control is the same thing as the sense of agency, I accept that the sense of agency is relevant for the agent's judgment of control. Therefore, the experiment does reveal something relevant about sense of agency, and it provides evidence for the claim that action selection is important for sense of agency.

11 It is noteworthy that the experimenters equate action selection to intention specification.

Based on the experiments, Haggard (2017) argues that preparation for action is also a crucial element for the sense of agency. When there is a high probability that the predicted outcome of the action will happen, intentional binding occurs even when the outcome does not occur (Moore and Haggard, 2008). One may conclude that more than one element, such as action selection, as well as the outcome prediction made by the comparator model, contribute to the sense of agency. Haggard (2017, p. 204) suggests that there may be a link between premotor signals of action selection and perceptual signals, which arise from body movement and its outcomes. Therefore, action selection, the match between motor intention and action feedback, and perception of the relation between action and outcome of action may contribute to give rise to the sense of agency. Each mechanism may contribute in its specific way, and the sense of agency helps the agent experience the action as self-caused (Haggard, 2017).

If the above interpretation of the data is correct, how does this description of neuro-cognitive mechanisms affect claims about the sense of agency? I have been discussing neuro-cognitive mechanisms, not the phenomenal aspect of the sense of agency. ATA may contend that none of the empirical information says anything about the latter and that their claims are about the agent's sense of agency, not about the mechanisms associated with it. The biological processes are one thing, and the phenomenology is a different thing. So, what relevance do these biological processes have for statements about the phenomenology of action?

I contend that the neuro-cognitive mechanisms that give rise to the sense of agency are relevant to ATA's claims, and one evidence for this is that tweaking the mechanism affects the sense of agency. Agents seem to feel it more or less strongly, depending on whether certain functions of the mechanism in action production are performed or not, or depending on the outcome of the action (Beck, Di Costa, Haggard, 2017; Imaizumi and Tanno, 2019). Since how acting feels is directly modulated by these mechanisms, empirical data from experiments involving these mechanisms are relevant to understand the sense of agency. Additionally, considering that the relevant functions of the mechanisms are, roughly, action selection and control, it is reasonable to conclude that the sense of agency is the sense/experience of action selection and control, or perhaps an even broader system involving correspondence with the agent's intention. However, its content may be too thin to make this explicit. I will defend in the following section that this is not compatible with how ATA advocates understand the sense of agency and its relation to their theories, and what this means for the claims made by ATA defenders.

4. The scientific account undermines claims made by ATA

4.1. Anti-reductionist Theories of Action

What I call ATA refers to theories that offer an account of human action characterized for not reducing the role the agent plays in the production of her action¹² (specifically in action initiation) to her mental states or events in their accounts of intentional action. In this sense, they oppose CTA, which, roughly, advances that the causal role the agent's relevant mental states and events play in the production of her action accounts for the action. ATA usually come in the form of Agent Causation or Non-causal Theory (volitionism).

Agent Causation presents different versions. In general, defenders of the theory (Chisholm, 1964; O'Connor, 2000; Clarke, 2010; Nida-Rümelin, 2007) claim that the agent directly causes¹³ her action in an irreducible way; typically, they defend that the causal relation involved in the production of human actions is substance causation (Clarke, 2010),¹⁴ instead of event causation.

There are also different versions of Non-causal Theories. Roughly, advocates of these theories defend that acts of will (willings or volitions)¹⁵ are basic mental actions of the agent, which are uncaused (Ginet, 1997; Lowe, 2008). The uncaused volition then causes the agent's bodily movements, and this is how the agent produces changes in the world. It is also an anti-reductionist theory because volitions are the exercise of the agent's power, or capacity, to will. The agent's *power* is not reduced in this view, and the volition *is* the agent's action (Lowe, 2008, p. 148).

I consider it fair to call these theories anti-reductionist because what characterizes them, from my point of view, is the claim that the agent's role in the production of her action cannot be reduced to her mental events or states playing a relevant role in the production of her action. The agent herself plays the central role in action production by originating (or bringing about) her

12 Although human agents often perform mental actions, I will not focus on mental actions here (except for volitions, see note 1). When I refer to actions henceforth, I mean overt intentional actions.

13 Direct causation relates to Agent Causation's understanding that the agent makes something happen (her action) in a causal understanding of making something happen, combined with the idea that she makes something happen directly; i.e., without the intermediation of any other causal step. One must keep in mind that direct causation is a term used in a conceptual framework that accepts substance causation (Chisholm, 1964). For more on substance causation, see note 14.

14 Roughly, the agent is an enduring substance that, as an agent, has the causal power to cause her actions; i.e., the causal relation is between the agent, as substance, and an internal event of hers. The agent's power is not reducible to her biological properties (O'Connor, 2000).

15 A mental action that is not caused by any previous mental event and this mental action causes the agent's movements. According to Carl Ginet (1997, p. 89) it has the phenomenal quality of feeling to the agent that she made the movement occur.

action.¹⁶ Alternatively, CTA proposes that the agent's mental states play the relevant causal role in the production of her action, thus providing a reductionist explanation of action production. In the sense that the agent's role is reduced to her mental events' causal role. CTA advocates may espouse different versions of CTA (Davidson, 1963; Mele, 1992, 2003; Enç, 2003; Aguilar, 2012; Shepherd, 2014); here, I will accept the theory's general proposal that the agent's mental events or states causally contribute to the production of her intentional action, which accounts for the action.

I will concentrate on three allegations made by Anti-reductionist Theories of Action (ATA from henceforth):

- (1) The sense of agency supports ATA;
- (2) ATA capture the human sense of agency;
- (3) CTA cannot capture (the content of) the human sense of agency.

I turn now to claims made by Non-causalists in relation to the sense of agency. They state that the sense of agency supports, or provides evidence, that the Non-causal account of actions should be favored. Lowe seems to consider that the sense of agency is a conscious sense, or awareness of the agent's willing: "I would, moreover, contend—I know somewhat controversially—that the phenomenology of voluntary action in fact supports non-causal theories, in that we are at least sometimes consciously aware of our willings as such" (Lowe, 2008, p. 154). The phenomenology of acting voluntarily encompasses the sense of agency and the sense of ownership. Lowe, however, is probably more concerned with the aspects of the phenomenology that characterize the sense of agency, because his theory is not about bodily movements, but about actions (this is the same for all other ATA discussed below). He believes that the said phenomenology, the sense of agency, is evidence of support for his theory about the agent's will (or volition). Therefore, Non-causalists advance the claim that (1) the sense of agency supports Non-causal Theories.

16 Not all theories I call ATA account for actions focusing only on the irreducible role of the agent; some theories are mixed. By mixed, I mean that they encompass reductionist bits in their accounts of action. Perhaps most ATA are mixed in some sense, but the ATA I am concerned with here are the ones that attribute action initiation to the agent's irreducible role, and claim that this part of their views is supported by, or accounts for, the sense of agency. An example of a mixed view that I will not be concerned with is Pamela Hieronymi's (2009) account of action. Her theory escapes my scope because, first, I am not aware that she makes any claims about the sense of agency, and most importantly, because in her mixed theory, the agent's irreducible role is not played by initiating action, but upstream, in practical reasoning (settling whether to A). Therefore, by the own theory's placement of the agent's irreducible role, it cannot be associated (support or be accounted by) the sense of agency. The sense of agency is the experience of the action as self-caused, not of answering the question of whether to A.

I broaden the scope of the claim to encompass ATA in general because, for instance, Bayne and Levy (2006) consider that Agent Causationists interpret the sense of agency in agent-causal terms; i.e., that agent causation can be experienced by the agent. They say that “as the term suggests, agent causal theorists adopt an explicitly causal conception of the phenomenology of being an agent” (Bayne and Levy, 2006, p. 56).¹⁷ Thus, allegedly, (1) the sense of agency supports ATA.

Considering claim (2), Timothy O’Connor states: “First and foremost (as I suggested at the outset), the agency theory [Agent Causation] is appealing because it captures the way we experience our own activity” (1995, p. 196). Posteriorly, O’Connor also claims:

(a) the content of the experience-in-acting of ordinary human agents involves a fairly inchoate sense of themselves as bringing about their actions and that (b) the reflective account that best captures this inchoate content is the agent-causal account (O’Connor, 2009, p. 34).

In the above passage, O’Connor accepts that Agent Causation Theory explains the human agents’ experience of producing intentional actions, since the agent herself originates her action by directly causing it. The suggestion that the agent directly causes her action allegedly explains the sense of agency. Therefore, he claims that (2) Agent Causation captures the human sense of agency. I generalize this claim to ATA, because if it is shown that it is not true for ATA, then it is also not true for Agent Causation.

Agent causationists also claim that CTA cannot account for the sense of agency as Agent Causation does. For instance, Nida-Rümelin (2007) claims that theories that do not take the phenomenology of action into account fail to provide a theory compatible with the everyday experience of human agents. Pereboom agrees with this objection to CTA (2015, p. 283), and Lowe believes that CTA is false to the phenomenology of action (2008, p. 189). Therefore, they seem to agree that (3) CTA cannot capture (the content of) the human sense of agency. In the next section, I will investigate the empirical data that enables evaluating whether these ATA theorists have reasonable grounds to make these claims.

¹⁷ Derk Pereboom (2015) agrees that the phenomenology of acting may make it seem that human agents cause their actions in accordance with Agent Causation’s explanation of this causation, and that CTA is insufficient to account for the phenomenology of the production of human actions.

4.2. *The sense of agency does not support ATA*

I will argue that there is not enough reason to believe ATA's claim because it is not evident that the empirical data lends support to ATA. Claim (1), in particular, seems to appeal to one's intuitions. It takes for granted that it is intuitive that agents have the experience/sense of irreducible origination of their actions; i.e., they claim that the sense of agency supports their theories of irreducible origination of action. Additionally, this is considered a reason for accepting ATA. It is unclear how the sense of agency allegedly supports ATA. I will call this judgment an intuition.¹⁸ It is as if from one's experiences/sensations or their contents, one had the intuition that the world should be explained in a certain way—i.e., under ATA—and, from this intuition, one concludes that the sense of agency supports ATA. Thus, ATA is true.

The problem with intuitions is that the debate is swayed to wherever the intuition of each philosopher involved in the debate turns (Knobe and Nichols, 2008; Deutsch, 2015). Additionally, treating the mental state of intuiting (or of considering a claim intuitive) as evidence for the content of the intuition—i.e., for what the intuition is about—is not adequate support in a philosophical argument (Deutsch, 2015). A further step is necessary: it needs to be shown that the sense of agency is the *phenomenal correlate* (the experience) of agent-causation or of the volition.¹⁹

Claim (2)—that ATA capture the sense of agency—sets off from theory and claims that it is the best account of the sense of agency. In order to evaluate the claim, it is necessary to understand what the sense of agency is, and how it comes about. Questions about the mechanisms that bring about the sense of agency and its production are empirical questions. The kind of theories defended by ATA are not theories intended to be empirical or to propose empirical hypotheses. Additionally, they may even eschew the idea that these claims should be compatible with the data that science brings into the discussion. Here, I will assume that the latter is not the case.

Nevertheless, the fact that the kind of claims ATA advocates make are not intended as dependent on empirical support does not render them immune to the issues raised by the empirical data about the sense of agency. Direct

18 I will not provide an account of intuitions; all I mean is that it is a judgment for which an argument has not been offered (which does not mean that an argument for it could not be offered).

19 Perhaps intuitions can be useful in cases that depend on social conventions; e.g., how English speakers conceive of a term, how they use language in certain situations, or what they mean (or presuppose) in their use of language. Nevertheless, these are distinct from the use of intuitions discussed here.

disagreement with our knowledge about the world (at the moment) is a good reason to be suspicious about these theories. Moreover, ATA theorists have taken a step further by making claims about something that is an empirical question: their theories capture the sense of agency, which can be studied empirically. Consequently, ATA should be compatible with the empirical data, and more so than competing theories. The data should offer corroborative evidence²⁰ for ATA in detriment of CTA.

If ATA captures the sense of agency, then it would be expected that the empirical data would match ATA. I will investigate some possible expectations. Although I do not mean to exhaust these possibilities, time and cause are critical aspects of empirical data,²¹ and it is crucial to ATA's claim what the data might inform about the content of the sense of agency. I discuss each, as well as content, in the following.

The temporal aspect of ATA would suggest that it is associated with triggering the action production or the muscle movement, so the sense of agency would arise at one of these moments when the motor command initiates the action production or just before the motor movement. As seen in the previous section, this is not the case; the underlying mechanisms of sense of agency are much more complex, and they are part of the action control processes. However, it may seem that the experiments and conclusions about action selection fluency and sense of agency favor ATA. I disagree for the following reasons. In the discussion about action selection, it is clear that the prime that antecedes action selection influences it significantly. Considering that ATA shun the idea that agent causation or the *will* (volition) can be led to act by the prime (a previous event), especially if a causal connection is suggested, then the role of the prime seems to show ATA's impossibility to account for the sense of agency. This is the case because the sense of agency is present even in cases in which the prime facilitates action selection, boosting the sense of agency.

Moreover, other processes give rise to the sense of agency, which do not come into place at the expected moments, such as the comparator model (Haggard, 2017) and the perception of the relation between the action and outcome (Wen; Haggard, 2020). If these processes give rise to the sense of agency, and ATA's claims about the sense of agency are not coherent with

20 By corroborative evidence I mean data that add reasons to believe in the theories' claims.

21 The local aspect of the evidence suggests no particular support or counterevidence to ATA, from my perspective. It is possible to conjecture that the area of the brain involved in agent causation or the will should be the area associated to selecting, planning or triggering action, but this is not made explicit by ATA. The sense of agency seems to rise from some of these mechanisms, or from their combination, but it is not limited to them. Mechanisms of action control are also important, as well as perception.

the facts about the sense of agency, then ATA is making claims that are not the case. One way ATA can be coherent with all these processes would be if it postulated an enduring cause of action, triggering action as well as playing a role in its control. Perhaps ATA would claim that the action plan and the goal it should accomplish are encompassed in the agent causation or the *will*. Even if this was coherent with the comparator model, it is not coherent with the perception of the relation between the action and its outcome, which can also give rise to the sense of agency. The perception is posterior to the action, so it does not take place at the moment that agent causation or the will would manifest to trigger action. Hence, it does not fit with the experiment described by Wen and Haggard (2020).

I must concede that experienced time often does not coincide with objective time (Libet, 1985; Haggard and Clark, 2003).²² ATA do not all clarify whether agent causation or the will/volition have a neuro correlate, but these are conceived as the agent's role in the production of the action; therefore, they are conceived as something of which the agent is conscious. And not just conscious of any content, but conscious as an experience of initiating action, this is precisely the phenomenological claims ATA makes. Could it be that the experience is delayed because objective time differs from experienced time? It is possible, however, as the experiments in section three show, the experience arises from control mechanisms of action execution that are at work during the action production, not from action initiation. One piece of evidence is that these mechanisms modulate the sense of agency. The modulation provides no reason to accept ATA's claims that associate the sense of agency with the will or agent causation. If Libet's (controversial) results are correct, subjects reported moving their finger before they moved (the M series), which could suggest that if the alleged experience of agent causation or the will should be anticipated, not delayed. ATA's claims' failure to match the sense of agency's timeline cannot be explained by such anticipation.

The causal aspect requires that if ATA is associated with the sense of agency, then the latter must spring from agent causation or the will, in the sense that it must originate from one of these. Again, it is hard for ATA to fit with accounts of the sense of agency that are coherent with experimental results. The latter show that the sense of agency can be associated with the perception of the relation between action and outcome. This kind of experiment suggests that the sense of agency can spring from the perceived relationship between the

22 I thank an anonymous referee for pointing out this important objection.

action and its outcomes—events in the world. However, what brings about the action in ATA—agent causation or the will—allegedly occurs even in cases in which the agent does not perceive the relation (the 60% interference condition) when she does not have the sense of agency. If ATA's account of how the agent brings about her action was associated with the sense of agency, then the agent should have a sense of agency even in the 60% interference condition because the action did occur.

The presented data seems enough to show that sense of agency and ATA explanation are independent and that ATA's account of how agents bring about their actions cannot be associated with the sense of agency. If it could, the sense of agency would be preserved in all interference conditions, because agent causation or the will allegedly bring about the agent's action in all conditions.

The content of the sense of agency is a difficult matter. Provided that it is such a thin phenomenology, it is hard to say what its content might be. One possibility is that it might reflect the function played by the sense of agency or of what the experience is. Another possibility is that the agent might derive the content from what she perceives when she has the sense of agency, or from inferences she makes based on these and her beliefs—something like what Pereboom (2015) calls an impure phenomenology. If the first possibility is correct, the content might be an experience of action as self-caused (that it was the agent's own action). If the second possibility is correct, then the content might be much more comprehensive; however, it would not indeed be the content of the sense of agency because it would be mixed with the agent's previous beliefs. For these reasons, it cannot lend substantial support to ATA.

I do not claim that agents are wrong about the sense of agency, i.e., their conscious experience. All I am saying is that experiences sometimes do not provide the best account of events. Even if it may seem to agents that the sense of agency is the experience of the will or agent causation, it is wise to question if it can be taken at face value. For instance, when I taste a glass of wine, I feel like I have a gustatory sensation. Nevertheless, the taste of wine is a complex sensation resulting from a combination of many perceptual inputs from different senses, like taste, touch, and smell (Smith, 2013). I do not intend to criticize ATA theories here. What I question are their claims about the sense of agency, claims that take the phenomenology at face value. A theory that makes claims about the sense of agency must be compatible with the neuro-cognitive information about the mechanisms and processes that go on throughout action production, which are associated with action selection and control.

Taking the phenomenology at face value can be dangerous because it might be that not every claim based on phenomenology is true. Someone suffering from phantom limb pain may claim that her amputated leg hurts. Now, we can take this at face value and believe that there is a phantom leg there. We may even accept a theory that postulates that amputated limbs remain ghostly attached to the body. Nonetheless, scientific explanation for why the person feels pain as if it were on her leg (that is no longer there) shows that it is not the leg that hurts—simplifying one possibility, the pain might occur due to reorganization of the nervous system. The scientific information does not show that there is no phantom limb there, but it makes the postulation of a phantom limb unnecessary because the phenomenology can be explained in a naturalistic way.

Summarizing, traditionally, it has been defended that agent causation and volitions play roles of initiating action in ATA—triggering the action. Agent Causationists claim that the agent directly causes her action. Therefore, the claims about the sense of agency supporting ATA should show that the experience is associated with the causation. I am not aware of any literature that makes clear the details of how direct causation works, but I assume that defenders of Agent Causation mean that direct causation triggers the brain process that initiates the production of the action.²³ The explanation above suggests that it is associated with more complex processes.

The issue is similar for volitions, perhaps even more puzzling. The non-causalist can claim that the agent's volition is the agent's mental action, which triggers the action results (bodily movements). I assume that the sense of agency would support the theory if it could be associated with the volition because the volition is the action. Nevertheless, I am not sure how the volitionist would explain that the sense of agency arises after the agent's volition. Considering that volitions *are* the agent's action, one would expect that the sense of agency should accompany the volition, which cannot be the case if the volition is a mental action that comes about prior to the workings of the control mechanisms that give rise to the sense of agency. The processes that produce the sense of agency would occur after the volition comes about; so, the non-causalist cannot maintain that her theory captures the sense of agency. Additionally, participants

23 Chisholm (1964) suggests that the agent may cause some event in her brain when she agent causes her action, say, moving her hand. Timothy O'Connor (2009) defends that agent causation brings about an intention that plays a role in the production of the movement. If I am correct, these instances of agent causation should be associated with the sense of agency.

in the experiment, for instance, willingly accept to follow the instructions of the experiments, but it is not clear if they have a sense of agency associated with such decisions.

I am assuming that both agent-causation and the volition somehow cause an event, or bring about a process, in the agent's brain. I consider that this is a punctual event instead of, for instance, a multiplicity of causations temporally following one another, and here is where ATA go awry. The sense of agency is not associated with a punctual event, but with a process that involves various events and does not end at action selection. It seems to continue even after the movement is over while the intended outcome is observed. If ATA defenders wish to claim that the sense of agency is the sense of the agent irreducibly producing her action, the role the agent plays in this production should be accompanied by the sense of agency, especially if the sense of agency is counted as support for these theories. Nevertheless, the story told by the mechanisms associated with the sense of agency does not fit with ATA's proposal. The lack of fit seems like a serious issue for ATA's claim: (2) ATA capture the human sense of agency.

On the other hand, CTA aims, in general, to be compatible with what science tells us about human actions. I believe that the empirical knowledge we have about the sense of agency favors CTA because it accepts that the agent's mental events play the relevant causal role in the production of action. Nevertheless, CTA does not postulate that the sense of agency needs to be associated with a specific part of the production of action. CTA accepts that monitoring and control mechanisms must be in place to guarantee the production of the relevant movement; for instance, it has been proposed that intentions play not only a triggering role but also roles of monitoring and controlling the production of action (Bratman, 1984; Mele, 1992). One possibility is that the agent's intention structures the working of the comparator model (Shepherd, 2014). For these reasons, CTA is not incompatible with the hypothesis that the sense of agency is associated with the neuro-cognitive mechanism of action selection, monitoring, control, and with the perception of the relation between the action and its outcome. The production of action is a series of causally related events extended in time, which is compatible with what we know about the sense of agency.

It does not matter which version of ATA is favored; they require that the sense of agency should be associated with the agent's alleged irreducible role in the production of her action in order for ATA to account for it, or at least I assume that this would be the most reasonable hypothesis for ATA. However, the sense of agency is a complex experience associated with action selection and

control over action execution. Therefore, I believe this is enough to conclude that ATA have no ground to claim that they capture the phenomenology of acting. The opposite may be true; compared to CTA, the empirical data about the sense of agency seems to favor CTA. This is an issue for ATA's claim 3.

A clarification is in order. I argued against ATA's claim that the sense of agency provides evidence for their theories about the agent's irreducible role in action initiation (see section 4.1.). I defend that CTA's account of action initiation, execution, and monitoring fits better with what we know about the sense of agency and its underlying mechanisms. However, the discussion here does not focus on other parts of the action's etiology, before action initiation and unassociated to the sense of agency, such as practical reasoning, nor does it focus on whether the agent can play an irreducible role in them. Therefore, it is more precise to claim that the sense of agency fits CTA's explanation focused only on action initiation, execution, and monitoring.²⁴ However, it cannot help us decide about reductionism or anti-reductionism in other parts of the action's etiology, such as practical reasoning (see note 16), because the sense of agency is not associated with these processes.

5. Possible objections

One possible objection is for ATA to claim that agent causation or volitions can be associated with action selection—upstream of the triggering of the motor movement—which some of the abovementioned experiments show that gives rise to the sense of agency. The objection is similar to objection (1), stated below, so the answer to (1) is also an answer to this objection.

In a similar line, ATA could restate their claims, making the case that the phenomenology they have been talking about is the experience of choosing, which accompanies deciding to act. Osaín Deery (2015) takes reports about the experience of choosing as testifying to its existence and relevance. The experience of choosing seems to suit ATA's theoretical commitments better, although it is not clear that the experience of choosing can support their views of agent causation or volitions. The experience of choosing seems to go well with their view of the indeterministic nature of human action—it is noteworthy, however, that this characteristic is closer to their claims about free will than to

24 Some might prefer to call this CTA*. Independently of how it is called, this is the way in which it is used here.

their claims about action initiation.²⁵ There are two ways in which the objection can be interpreted:

- (1) The sense of choice can be associated with action selection mechanisms, so it contributes to the sense of agency.
- (2) The sense of choice is independent of the sense of agency.

Accepting (1), ATA could claim that action selection is the manifestation of agent causation or volition and that it elicits the sense that the action was self-caused, i.e., caused by the agent. If this were correct, it would be enough to evidence ATA's accounts of action initiation. I concede that in bringing about action, the agent may select (choose) among possible actions. However, this is precisely the point; ATA's claims are not about merely selecting action, they are about action initiation. Additionally, other mechanisms also elicit the sense of agency, such as the observation of the consequences of the action, so it would be difficult for ATA to explain how that might be the case if, according to the theories, the sense of agency arises from the agent playing her irreducible role in action initiation. How could other mechanisms in the production of action also elicit the sense of agency? The most parsimonious explanation is to deny ATA's claims and accept that the sense of agency is not an experience of this kind of action initiation but of various control mechanisms in the production of action, the comparator model, observation of consequences of the action, and action selection are control mechanisms in the production of the action.

As explained in section three, action selection contributes to the sense of agency, but it says nothing about whether agent causation or a volition initiates action. These mechanisms show no evidence in favor of agent causation or volitions. This is the case because looking at the mechanism that elicits the anemic phenomenology in question, one learns how it contributes to the sense of agency through fluidity or likelihood of accomplishing the intended result, and these do not provide any evidence in favor of ATA's claims.

In any case, I think ATA would probably stick with (2) and claim that this is a different experience. I see no problem in conceding that the sense of choice is an experience in its own right, independent of the sense of agency.

It is not clear with which mechanism it would be associated, but Deery convincingly argues that it is a combination of prospection (of alternative possibilities) and causal modeling of possible consequences of each choice

25 Deery (2015) provides a compelling discussion about claims about free will related with sense of choice.

(screening off any other causal antecedent to focus on the possible choice's consequence). If this is correct, then it is prospecting alternative possibilities and calculating their possible consequences that give rise to the sense of choice. The point is that the experience of choice and its underlying mechanisms are mute about how the action is initiated. Nothing about it evidences agent causation. The volitionist is in a more comfortable position, I believe, because she can claim that the decision is the emergence of the volition, whatever comes after are action results, and she is not committed to explaining those. This is fine, but the crucial part of volitionism is its proposal that volitions are uncaused actions. It is unclear how the experience of choosing might provide evidence of the lack of antecedent causation. Even if the experience of choice might seem indeterminist, if Deery is correct, the indeterminist appearance of the experience of choice is illusory because of the way the causal model works, screening off all antecedent causes to focus only on possible choices and their consequences. Therefore, I do not think that focusing on the experience of choosing helps ATA's case.

Conclusion

In the present investigation, I challenged ATA advocates' claim that what I have been calling the *sense of agency* lends support, especially for ATA, in detriment of CTA. To do so, I discussed what it means to say that the sense of agency supports or is captured by ATA. Considering that the conclusion was that these claims have empirical implications and that there is empirical information available, the next step was an overview of the empirical information. Finally, I argued that, when thoroughly inspected, the sense of agency does not support ATA; it seems to support CTA.

Acknowledgments

I am grateful to Grupo de Escrita de Mulheres na Filosofia (GEMF), Josh Turkewitz, and Marcelo Fischborn for comments on a previous version of this paper. I would also like to thank the anonymous reviewers for their comments and suggestions.

References

- AGUILAR, J. "Basic Causal Deviance, Action Repertoires, and Reliability". *Philosophical Issues*, 22(01), 2012, pp. 1-19.
- BARLAS, Z., OBHI, S. "Freedom, Choice, and the Sense of Agency". *Frontiers in Human Neuroscience*, 7(514), 2013, pp. 1-7.
- BAYNE, T., LEVY, N. "The Feeling of Doing: Deconstructing the Phenomenology of Agency". In: N. Sebanz, W. Prinz (ed.). *Disorders of Volition*. Cambridge, MA: MIT Press, 2006. pp. 49-68.
- BECK, B., DI COSTA, S., HAGGARD, P. "Having Control Over The External World Increases The Implicit Sense of Agency". *Cognition*, 162(201705), 2017, pp. 54-60.
- BRATMAN, M. "Two Faces of Intention". *The Philosophical Review*, 93(3), 1984, pp. 375-405.
- CHAMBON, V., HAGGARD, P. "Sense of Control Depends on Fluency of Action Selection, Not Motor Performance". *Cognition*, 125(3), 2012, pp. 441-451.
- CHISHOLM, Roderick. "Human Freedom and the Self". In: G. Watson (ed.). *Free will*. Oxford: Oxford University Press, 1964. pp. 24-35.
- CLARKE, Randolph. "Agent Causation". In: T. O'Connor, C. Sandis (eds.). *A Companion to the Philosophy of Action*. Oxford: Wiley-Blackwell, 2010. pp. 218-226.
- DAVIDSON, D. "Actions, Reasons and Causes". *Journal of Philosophy*, 60(23), 1963, pp. 685-700.
- DEERY, O. "Why People Believe in Indeterminist Free Will". *Philosophical Studies*, 172(8), 2015, pp. 2033-2054.
- DEUTSCH, M. "The Myth of the Intuitive: Experimental Philosophy and Philosophical Method". Oxford, MA: MIT Press, 2015.
- ENÇ, B. "How We Act: Causes, Reasons and Intentions". Oxford: Oxford University Press, 2003.
- FEINBERG, I. "Efference Copy and Corollary Discharge: Implications for Thinking and Its Disorders". *Schizophrenia Bulletin*, 4(4), 1978, pp. 636-640.
- FRITH, C. "Explaining Delusions of Control: The Comparator Model 20 Years On". *Consciousness and Cognition*, 21(1), 2012, pp. 52-54.
- GALLAGHER, S. "Multiple Aspects in the Sense of Agency". *New Ideas in Psychology*, 30(1), 2012, pp. 15-31.
- GINET, C. "Freedom, Responsibility, and Agency". *The Journal of Ethics*, 1(1), 1997, pp. 85-98.
- GRÜNBAUM, T. "The Feeling of Agency Hypothesis: A Critique". *Synthese*, 192(10), 2015, pp. 3313-3337.
- HAGGARD, P. "Sense of Agency in the Human Brain". *Nature Reviews Neuroscience*, 18(4), 2017, pp. 196-207.
- HAGGARD, P., CLARK, S. "Intentional Action: Conscious Experience and Neural Prediction". *Consciousness and Cognition*, 12(4), 2003, pp. 695-707.
- HIERONYMI, P. "The Will as Reason". *Philosophical Perspectives*, 23(1), 2009, pp. 201-220.

- IMAZUMI, S., TANNO, Y. "Intentional Binding Coincides with Explicit Sense of Agency". *Consciousness and Cognition*, 67, 2019, pp. 1-15.
- KNOBE, J., NICHOLS, S. "Experimental Philosophy". Oxford, New York: Oxford University Press, 2008.
- LIBET, B. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action". *Behavioral and Brain Sciences*, 8(04), 1985 pp. 529.
- LOWE, E. J. "Personal Agency: The Metaphysics of Mind and Action". New York: Oxford University Press, 2008.
- MELE, A. "Motivation and Agency". New York: Oxford University Press, 2003.
- _____. "Springs of Action: Understanding Intentional Behavior". New York: Oxford University Press, 1992.
- MOORE, J., HAGGARD, P. "Awareness of Action: Inference and Prediction". *Consciousness and Cognition*, 17(1), 2008, pp. 136-144.
- MYLOPOULOS, M. "Agentive Awareness Is Not Sensory Awareness". *Philosophical Studies*, 172(3), 2015, pp. 761-780.
- MYLOPOULOS, Myrto, SHEPHERD, Joshua. "Agentive Phenomenology". In: U. Kriegel (ed.). *Oxford Handbook of the Philosophy of Consciousness*. Oxford: Oxford University Press, Forthcoming.
- NIDA-RÜMELIN, M. "Doing and Subject Causation". *Erkenntnis*, 67(2), 2007, pp. 255-272.
- O'CONNOR, T. "Agent Causal Power". In: T. Handfield (ed.). *Dispositions and Causes*. Oxford: Oxford University Press, 2009. pp. 189-214.
- _____. "Persons and Causes: The Metaphysics of Freewill". New York: Oxford University Press, 2000.
- _____. "Agent Causation". In: T. O'Connor (ed.). *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press, 1995. pp. 173-200.
- PACHERIE, E. "The Phenomenology of Action: A Conceptual Framework". *Cognition*, 107(1), 2008, pp. 179-217.
- PEREBOOM, D. "The Phenomenology of Agency and Deterministic Agent Causation". In: H. Pedersen, M. Altman (eds.). *Horizons and Authenticity in Phenomenology, Existentialism, and Moral Psychology*. New York: Springer, 2015. pp. 277-294.
- SCHROETER, F. "Endorsement and Autonomous Agency". *Philosophy and Philosophical Research*, 69(3), 2004, pp. 633-659.
- SHEPHERD, J. "The Contours of Control". *Philosophical Studies*, 170(3), 2014, pp. 395-411.
- SMITH, B. "The Nature of Sensory Experience: The Case of Taste and Tasting". *Phenomenology and Mind Online Journal*, 2013, pp. 292-313.
- SYNOFZIK, M., VOSGERAU, G., NEWEN, A. "I Move, Therefore I Am: A New Theoretical Framework to Investigate Agency and Ownership". *Consciousness and Cognition*, 17(2), 2008, pp. 411-424.
- WEGNER, D. "The Illusion of Conscious Will". Cambridge: The MIT Press, 2002.

WEN, W., HAGGARD, P. “Prediction Error and Regularity Detection Underlie Two Dissociable Mechanisms for Computing the Sense of Agency”. *Cognition*, 195, 2020, pp. 104074.

WENKE, D., FLEMING, S., HAGGARD, P. “Subliminal Priming of Actions Influences Sense of Control Over Effects of Action”. *Cognition*, 115(1), 2010, pp. 26-38.

WOLPERT, D. M. “Computational Approaches to Motor Control”. *Trends in Cognitive Sciences*, 1(6), 1997, pp. 209-216.

