

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Hoerl, Christoph, Lorimer, Sara, McCormack, Teresa, Lagnado, David, Blakey, Emma, Tecwyn, Emma C. and Buehner, Marc J. (2020). Temporal binding, causation and agency: developing a new theoretical framework. *Cognitive Science*.

Publishers page:

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

## **Temporal binding, causation and agency: Developing a new theoretical framework**

Christoph Hoerl<sup>1†</sup>, Sara Lorimer<sup>2</sup>, Teresa McCormack<sup>2</sup>, David A. Lagnado<sup>3</sup>, Emma Blakey<sup>4</sup>, Emma C. Tecwyn<sup>5</sup>, & Marc J. Buehner<sup>6</sup>

<sup>1</sup>University of Warwick, Department of Philosophy

<sup>2</sup>Queen's University Belfast, School of Psychology

<sup>3</sup>University College London, Department of Psychology

<sup>4</sup>The University of Sheffield, Department of Psychology

<sup>5</sup>Birmingham City University, School of Social Sciences <sup>6</sup>

Cardiff University, School of Psychology

† Corresponding author: Department of Philosophy, University of Warwick, Coventry CV4 7AL, United Kingdom. [c.hoerl@warwick.ac.uk](mailto:c.hoerl@warwick.ac.uk) ++44 24 7652 2476

## **Acknowledgements**

For comments and suggestions, we are grateful to John A. Dewey and four other reviewers for this journal. This research was supported by a Leverhulme Trust Research Project Grant (RPG-2015-267).

## **Temporal binding, causation and agency: Developing a new theoretical framework**

### **Abstract**

In temporal binding, the temporal interval between one event and another, occurring some time later, is subjectively compressed. We discuss two ways in which temporal binding has been conceptualized. In studies showing temporal binding between a voluntary action and its causal consequences, such binding is typically interpreted as providing a measure of an implicit or pre-reflective “sense of agency”. However, temporal binding has also been observed in contexts not involving voluntary action, but only the passive observation of a cause-effect sequence. In those contexts, it has been interpreted as a top-down effect on perception reflecting a belief in causality. These two views need not be in conflict with one another, if one thinks of them as concerning two separate mechanisms through which temporal binding can occur. In this paper, we explore an alternative possibility: that there is a unitary way of explaining temporal binding both within and outside the context of voluntary action as a top-down effect on perception reflecting a belief in causality. Any such explanation needs to account for ways in which agency, and factors connected with agency, have been shown to affect the strength of temporal binding. We show that principles of causal inference and causal selection already familiar from the literature on causal learning have the potential to explain why the strength of people’s causal beliefs can be affected by the extent to which they are themselves actively involved in bringing about events, thus in turn affecting binding.

**Keywords:** temporal binding; intentional binding; sense of agency; voluntary action; agency attribution; causality; causal inference

## **Temporal binding, causation and agency: Developing a new theoretical framework**

### **1. Introduction**

The subjective timing of events has been shown to be affected by a wide variety of factors.

In this paper, we focus specifically on two types of mechanisms that have been postulated to explain instances of *temporal binding*, i.e., instances in which the interval between one event and another, occurring some time later, is subjectively compressed.

A great deal of research on temporal binding has studied such binding in the context of voluntary action. A typical study will, for instance, involve participants pressing a button, which is then followed by a tone, with the finding being that the button press is judged to occur later and the tone is judged to occur earlier when compared to a set of baseline temporal judgements obtained by asking participants to time each of the two types of events occurring in isolation. Following Haggard, Clark, and Kalogeras (2002), it has become common to refer to binding found in the context of such studies as ‘intentional binding’, with the implied idea being that it occurs because the two events that are bound together are a voluntary action and its causal consequence. More specifically, Haggard et al. (2002) hypothesized that such intentional binding is reflective of a cognitive mechanism that allows individuals to experience a sense of their own agency, and they therefore argued that “the perceived time of intentional actions and of their sensory consequences [can be used] as a means to study consciousness of action” (Haggard et al., 2002, p. 382). This idea, that temporal binding can serve as a measure of a ‘sense of agency’, has since been adopted by a wide range of researchers (see, e.g., Barlas, Hockley, & Obhi, 2018; Demanet, Muhle-Karbe, Lynn, Blotenberg, & Brass, 2013; Moore, Wegner, & Haggard, 2009).

A quite different way of conceptualizing temporal binding has emerged in the context of some more recent studies that have demonstrated instances of temporal binding also

occurring outside of the context of voluntary action, for instance when participants passively observed another person's button press followed by a tone (Poonian, McFadyen, Ogden, & Cunnington, 2015), when an experimenter pressed down the participant's finger to initiate the tone (Borhani, Beck, & Haggard, 2017; see also Buehner, 2015), or when the button was simply depressed by a machine (Buehner, 2012). It has been hypothesized that instances of temporal binding such as this are a top-down effect on perception reflecting a belief in a cause-effect relation between the two observed events. In this context, researchers have also drawn parallels between temporal binding and the phenomenon of causal re-ordering (Bechlivanidis & Lagnado, 2013; 2016), in which participants subjectively invert the order in which two events happen if they are led to believe the later event to be the cause of the earlier one. Both phenomena – temporal binding and causal reordering – have been connected with basic Humean assumptions known to inform causal learning: the assumption that causes are temporally contiguous with their effects, and the assumption that causes precede their effects, respectively. More specifically, it has been suggested that temporal binding and causal reordering are brought about by processes that align the subjective timing of two causally related events with those basic Humean assumptions about causality – by making the events appear more contiguous with each other in the case of temporal binding, or by reordering so that the presumed cause is perceived as preceding the presumed effect (Bechlivanidis & Lagnado, 2016; Blakey et al., 2019; Buehner, 2012; Vuorre, 2017).

The series of more recent findings demonstrating that temporal binding occurs even in the absence of agency (e.g., Borhani et al., 2017; Buehner, 2012; Suzuki, Lush, Seth, and Roseboom, 2019) seems to force the conclusion that there are at least some circumstances in which temporal binding occurs as a top-down effect reflecting a belief in causality, rather than as the product of a sense of agency. How should we think of the relationship between explanations of temporal binding as a measure of a sense of agency and explanations of

temporal binding as a top-down effect reflecting a belief in causality? Clearly, the two types of explanation need not be in competition with one another, as they might simply single out two separate mechanisms by which binding can occur. If so, this might be thought to provide one possible explanation as to why studies that have compared a condition involving voluntary action with one containing only the observation of a cause-effect sequence have typically found stronger temporal binding in the voluntary action condition (Borhani, Beck, & Haggard, 2017; though see also Suzuki, Lush, Seth, & Roseboom, 2019) – the idea being that both mechanisms are operative in the case of an event brought about by a voluntary action, leading to an additive effect. We can call this general approach a two mechanisms account.

In this paper, though, we wish to explore an alternative possibility to a two mechanisms account: that there is a unitary way of explaining temporal binding both within and outside the context of voluntary action as a top-down effect on perception reflecting a belief in causality. Clearly, key to providing such a unitary explanation is that it, too, needs to be able to explain why temporal binding has often been found to be stronger in the context of voluntary action. However, as we will argue, even on an account that explains all instances of temporal binding as top-down effect of a belief in causality there are reasons to expect that agency can have a modulating effect on temporal binding and that temporal binding can also be affected by other factors that specifically affect participants' judgement about the degree to which they are themselves actively involved in bringing about an outcome.

Thus, a crucial part of our argument will be that our proposed unitary account, even though it construes temporal binding as a top-down influence of a belief in causality, nevertheless also has the resources to explain why agency can have an influence on the strength of binding. It will therefore be important to make clear just how this account differs from existing ones invoking the idea of a 'sense of agency'. This is why, in the next section,

we review such existing accounts, clarifying in particular the specific notion of a sense of agency in play in them and how exactly that notion has been invoked to explain experimental findings. In section 3, we then turn to the question of the relationship between accounts explaining temporal binding in terms of the notion of a sense of agency and accounts explaining it in terms of a top-down influence of causal belief on perception, and give reasons as to why this relationship deserves to be given more attention than it has received so far. Finally, in section 4, we develop our own theoretical framework for explaining temporal binding, which generalizes the idea that binding involves a top-down influence of causal belief on perception to also account for cases of temporal binding involving voluntary action. In particular, we show how existing principles of causal structure induction and causal selection, familiar from the literature on causal learning, can be invoked to explain how, even on such an account, agency might have a distinctive influence on temporal binding.

## **2. Temporal binding as a measure of a ‘pre-reflective sense of agency’**

Research on temporal binding started with a study by Haggard et al. (2002) in which temporal binding between a voluntary movement and its causal consequences was observed. To provide a brief sketch of the study by Haggard et al. (2002): Participants were required to watch a clock face on which a clock hand rotated and report the position of the clock hand when a particular event occurred. There were three experimental conditions, which were tested in a within-subject design: a voluntary action condition (in which the participants pressed a button), an involuntary movement condition (in which TMS caused the participants’ hand to twitch), and a sensory only condition (in which participants heard the click of a sham TMS). In each of these conditions a tone followed the relevant event after a 250ms delay. In half the trials in each condition participants judged the onset time of the first event (the button press, the TMS induced muscle twitch, or the click of the sham TMS), and

in the remaining trials they judged the onset time of the tone that followed the first event. These temporal judgements were then compared to a set of baseline temporal judgements made by each participant, obtained by asking participants to judge the onset time of each of the aforementioned events when they occurred in isolation. The researchers then calculated the magnitude of the perceptual shift that occurred as a result of the nature of each condition by subtracting the mean of the perceived onset time of a particular event at baseline from the mean perceived onset time of the same event when it occurred in an experimental condition. It was found that the voluntary actions and their consequent tones were perceived as occurring closer together in time. Specifically, when voluntary actions were followed by a tone, the action was perceived as occurring later compared with voluntary actions not followed by a tone, and the tone was perceived as occurring earlier compared with a tone not preceded by a voluntary action. By contrast, the opposite pattern was observed when the two events were a TMS-induced muscle twitch and a tone. That is, the muscle twitch was perceived as occurring earlier and the tone as occurring later, when compared with events of each type occurring in isolation. In the case of sham TMS clicks followed by a tone, no shift in the perception of the onset time of either event was observed when compared with each occurring on its own.

### **2.1. The idea of a ‘pre-reflective sense of agency’**

Since its publication, Haggard et al.’s (2002) study has attracted much interest and has inspired many other studies aimed at exploring aspects of agency that have used versions of a binding paradigm (e.g., Aarts et al., 2012; Engbert, Wohlschläger, & Haggard, 2008; Obhi & Hall, 2011). In general, researchers have followed Haggard et al. (2002) in assuming that temporal binding (at least as found in the context of voluntary action) reflects a sense of agency. In particular, it has become common to speak of temporal binding as an “implicit

measure of agency” as contrasted with the “explicit measure” provided by individuals’ actual verbal self-attributions of agency (Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Moore et al., 2009). Sometimes, the contrast is also described as one between temporal binding as a “measure of an implicit sense of agency” vs. verbal self-attribution as a “measure of an explicit sense of agency” (Beck, Di Costa, & Haggard, 2017; Kumar & Srinivasan, 2013).

As these latter descriptions bring out, the idea of a sense of agency itself has been conceptualized in quite a specific way in this context, and it is therefore important to distinguish between two quite different ways in which the term “sense of agency” can be understood. Sometimes it is used in a very broad sense, to designate all aspects of the awareness agents have of their own actions and that can underpin explicit judgements of agency (see, e.g., David, Newen, & Vogeley, 2008). The phenomenology of agency, thus broadly understood, encompasses a variety of different factors that can be used to make explicit judgements of agency, including cognitive factors such as the awareness of one’s own conscious intentions, and factors not exclusive to the carrying out of voluntary actions such as one’s proprioceptive awareness of one’s own bodily movements, or particular sensations of pressure, resistance, or tension.

By contrast, when researchers speak of temporal binding as a measure of a sense of agency, what they have in mind is that, within this overall phenomenology of agency, it is important to recognize a distinctive, basic feeling component characteristic specifically of voluntary actions. The idea has been that there is an “everyday” feeling of agency which, although a crucial aspect of the phenomenology of agency, is a background feeling that escapes attempts to make it the object of direct conscious introspection and may therefore not be reflected accurately in explicit judgements of agency. It is therefore also often referred to as a *pre-reflective* sense of agency. Haggard and Eitam (2015, p. xii), for instance, write that this purported pre-reflective sense of agency “more resembles a phenomenal background than

a clear and distinct perceptual experience”. Alternative descriptions refer to it as “non-conceptual and sub-personal” (Kühn, Brass, & Haggard, 2013, p. 1936) or a “background buzz of control” (Moore, 2016, p. 1).

Endorsing the idea of such a pre-reflective sense of agency Obhi and Hall (2011) argue that “[t]here is often a mismatch between the sense of agency as it is experienced in life (i.e. as a feeling that we do not dwell on) and how it is assessed in the laboratory... [E]xplicit judgments of agency, although important in their own right, are foreign to the low level, but essential agency we experience in life. Thus, there is a very real need to derive a measure of agency which does not depend on subjective report and which somehow captures the low level, pre-reflective nature of many agentic experiences” (Obhi & Hall, 2011, pp. 655f.). Thus, their argument is that temporal binding is an important way to measure sense of agency because the feeling in question is “phenomenally thin” (Haggard, 2005) or “elusive” (Kumar & Srinivasan, 2013), making direct measurement difficult or unreliable (see also Moore, 2016). A further line of argument is that explicit judgments of agency may be subject to biases, distortion, or social desirability effects in a way that the temporal binding measure is not (Haggard, 2017; Moore & Fletcher, 2012).

How exactly are we to think of the connection between a pre-reflective sense of agency, understood along the lines just sketched, and temporal binding? In the remainder of this section, we seek to provide a characterization of the general sort of theoretical framework that we take to be in play in existing accounts that connect temporal binding with the idea of a pre-reflective sense of agency. Note that, in doing so, we are not claiming that we have perfectly captured any one specific theoretical position. Indeed, part of our motivation in writing this paper is that we have found it quite difficult to gauge exactly how researchers have understood temporal binding and the nature of the processes that underpin it. Thus, we see it as valuable to try to get on the table a plausible account that embodies some key claims

that have been made by researchers in this area, which can then be compared with the alternative model that we will develop later in this paper.

## 2.2. Elements of the existing theoretical framework

Figure 1 depicts the key elements of the type of explanatory model that we take to be in play when researchers construe temporal binding as providing a way to measure pre-reflective sense of agency. The two elements in the top row – “Explicit judgments of agency” and “Temporal binding” – refer to two variables that are directly measured in some of the studies we draw on (though the relevant studies also include ones that only involve measuring temporal binding). For “Explicit judgments of agency”, this is done, for instance, by asking participants whether or to what extent they thought or felt that it was them who had produced a certain outcome. For “Temporal binding”, the variable represents the degree to which participants’ judgments about the onset times of an intentional bodily movement of theirs and

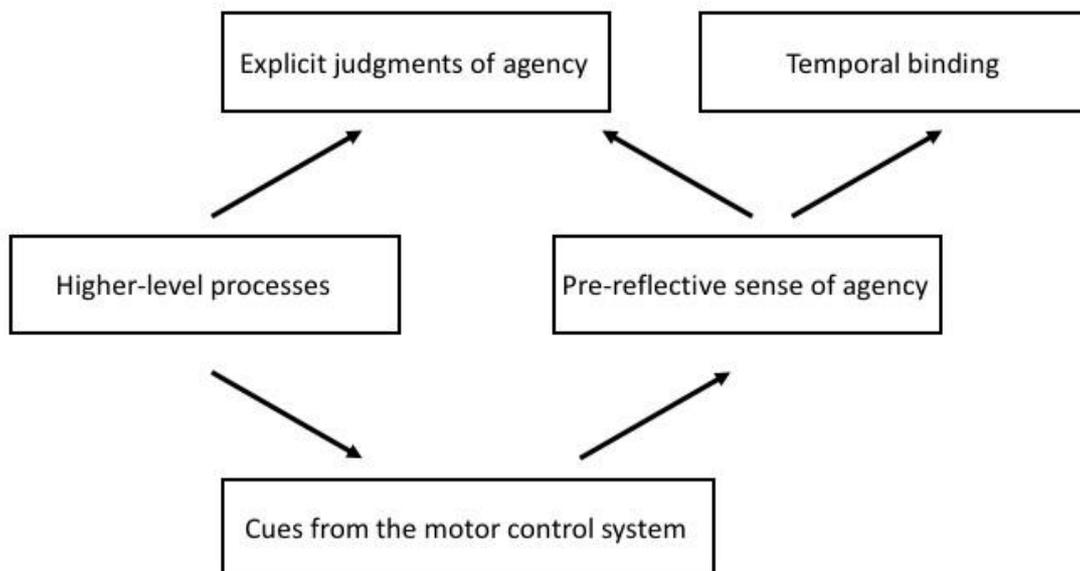


Figure 1:

A proposed characterization of the theoretical framework involved in explanations of temporal binding that construe it as a measure of a pre-reflective sense of agency

a subsequent distal event, or about the length of the interval between them, indicate a perceptual shift of these two events towards each other when compared with some suitable baseline set of events not involving intentional action.

The first thing to note, and of central concern to this paper, is that, as mentioned previously, the model construes temporal binding as a measure of a pre-reflective sense of agency. That is to say, it is assumed that temporal binding can serve as a measure of a specific feeling or experience – a particular aspect of the phenomenology of acting – that corresponds to the degree to which the individual controls or is responsible for their own actions and the consequences of those actions, but that does not require conscious reflection. Indeed, as we have said, researchers have argued that because it is a “background” feeling, it can elude attempts to make it the direct object of conscious introspection.

The model, as described, interprets the sense in which temporal binding is assumed to provide a measure of a pre-reflective sense of agency in terms of the idea that temporal binding is an expression of such a sense of agency, such that the presence of the former can serve as evidence of the presence of the latter. However, we note that the relationship between the two is also sometimes described in a way that implies a much tighter connection between them: i.e. that temporal binding is actually itself integral to, and part of what generates, the pre-reflective sense of agency (Aarts et al., 2005, p. 455; Haggard et al., 2002; Tsakiris & Haggard, 2003, p. 439; see also Mylopoulos, 2012, p. 111, on this issue). However, the difference between this more specific view and the general idea of temporal binding as providing a measure of a pre-reflective sense of agency will not be relevant to our argument in what follows, and we will therefore set it aside for our purposes.

What processes are assumed to underpin the pre-reflective sense of agency? In this model it is assumed that the processes in question are those of the motor control system itself. Note, though, that these may be processes involved in a number of aspects of the motor

control system at different stages of voluntary action. Historically, the emphasis has been on processes that occur after the action and its outcome have occurred – specifically, on the output of a comparator that compares the actual sensory consequences of an action with the predicted consequences, with it being assumed that people feel agency over actions because predicted and actual consequences match (Blakemore, Wolpert, & Frith, 2002). However, more recent expositions suggest that some aspects of action selection processes and the monitoring of such processes *in advance* of acting also serve to underpin sense of agency (Borhani et al., 2017; Haggard, 2017; Sidarus, Vuorre, & Haggard, 2017). In this model, we group these under the processes associated with the motor control system, although we recognize that they are distinct from those postulated by comparator models of the sense of agency.

What the hypothesized set of processes giving rise to the pre-reflective sense of agency have in common is that they are assumed to be sub-personal, not accessible to conscious awareness, and internal to the motor control system (see, e.g., Chambon, Sidarus, & Haggard, 2014; Synofzik, Vosgerau, & Newen, 2008; Wenke, Fleming, & Haggard, 2010). That is, the pre-reflective sense of agency, as an aspect of the phenomenology of acting, is conceived of as being the product of a set of processes that are not themselves conscious.

Apart from giving rise to or involving temporal binding, the pre-reflective sense of agency is also assumed to be what typically underlies explicit verbal self-ascriptions of agency. However, such explicit judgments are also assumed to be influenced by a different set of cues: a variety of processes that we have grouped together under the heading of “higher-level processes” that can affect explicit judgements of agency via a different pathway not involving the pre-reflective sense of agency. This is to explain why, although some studies find correlations between explicit judgments of agency and temporal binding (e.g., Barlas et al., 2018; Berberian et al., 2012; Caspar, Christensen, Cleeremans, & Haggard,

2016; Ebert & Wegner, 2010; Kawabe, Roseboom, & Nishida, 2013; Kumar & Srinivasan, 2013), there are also findings showing that the two measures can dissociate (e.g., Dewey & Knoblich, 2014; Lynn, Muhle-Karbe, Aarts, & Brass, 2014; Moore, Middleton, Haggard, & Fletcher, 2012; Saito, Takahata, Murai, & Takahashi, 2015; Wen, Yamashita, & Asama, 2015a).

The relevant processes hypothesized to affect explicit judgements of agency via a pathway not involving the pre-reflective sense of agency are higher-level processes in the sense that they are assumed to be cognitively based and can be conscious and/or inferential in nature, and we refer to them by the generic label “higher-level processes” because the processes in question form a diverse group. Explicit judgments of agency are known to be affected by a wide variety of factors external to the motor control system. Such factors include, for instance, beliefs about the mechanism by which the outcome is produced and whether there are any plausible alternative causes (Wegner, 2003). But there are also other types of influences on explicit judgments of agency. For example, they have been shown to be subject to self-serving biases, such that people are more likely to attribute positive or pleasant outcomes to their own actions (e.g., Barlas et al., 2018; Barlas & Obhi, 2013; Wegner, 2003) and judge that they are more in control following successful performance (Wen, Yamashita, & Asama, 2015b). We will discuss some further factors influencing explicit judgments of agency below when considering variables that have been shown to affect both explicit judgments of agency and temporal binding and how the model conceives of the pathways through which those variables affect both of those measures.

Current theories of explicit judgments of agency typically assume that people use a variety of cues when making such judgments, and that the purported pre-reflective sense of agency serves as just one cue amongst many (Moore & Fletcher, 2012; Synofzik et al., 2008). It is assumed that cues may be weighted differentially in a situation-dependent way that

reflects the reliability of different cue types (Moore & Fletcher, 2012). Synofzik et al. argue that in many circumstances in everyday life, if asked to make an explicit judgment of agency, people would take their pre-reflective sense of agency “at face value” and base judgments on this feeling. However, in more complex circumstances – for example, when there is more than one agent and there is ambiguity over who is responsible for an outcome – cues stemming from higher-level processes may be given greater weight, and may potentially override the pre-reflective sense of agency. The model in Figure 1 assumes that both cue types feed into explicit judgments of agency, although it is silent as to how exactly such cues are weighted to produce explicit judgments.

Finally, in the model it is assumed that higher-level cognitive processes can operate in a top-down way on processes grouped here as motor control processes (see, for instance, Desantis, Roussel, & Waszak, 2011). This is a means of explaining findings that indicate that temporal binding, too, can be affected by factors external to the motor control system. As with explicit judgments of agency, Yoshie and Haggard (2013) found that temporal binding was moderated by whether the outcomes resulting from actions were positive or negative (though see Barlas et al., 2018; Moreton, Callan, & Hughes, 2017). Manipulations that influence whether participants believe they are responsible for an action, such as varying whether actions are freely chosen versus carried out under instruction, coercion, or hypnosis also have an impact on temporal binding (Borhani et al., 2017; Caspar et al., 2016; Lush et al., 2017). Furthermore, when participants do actually cause an outcome but believe it is a result of another person’s actions, temporal binding is reduced or abolished (Desantis, Roussel, & Waszak, 2011). On the assumptions that (i) temporal binding is a measure of a pre-reflective sense of agency, and (ii) the pre-reflective sense of agency stems from the operation of motor control processes, then there must be some top-down influence on how those processes operate from the higher-level processes.

We note that there is an alternative possible explanation of these sorts of findings, which is that there are also mechanisms by which higher-level processes can influence temporal binding directly, rather than through having a top-down effect on motor control processes. For our purposes we need not rule out this possibility. Indeed, we will try to flesh out a version of this idea towards the end of this paper. Combining it with the broad type of theoretical framework depicted in Figure 1, though, would make it more difficult to hold on to the idea that temporal binding is a more direct way of measuring the experience of voluntary action than explicit judgments of agency (Moore & Obhi, 2012), because it would mean that both measures are affected by exactly the same set of factors. Some of the more specific ways in which the relevant research has been interpreted also seem to turn on the idea that higher-level processes influence temporal binding via a top-down influence on motor control processes. For instance, Yoshie and Haggard (2017) conclude from the fact that the influence of outcome valence on temporal binding depends on the predictability of the outcome valence that “a special link between the brain systems for computing valence and reward, and the brain systems for predictive control of action” (ibid., p. 1) is responsible for the effect. Similarly, Lush et al. (2017) conclude from the fact that hypnosis affects temporal binding that hypnosis changes the experience that people have when they act as well as influencing their reports about how voluntary their actions feel. The authors explicitly contrast this with the alternative hypothesis that the experience of actions carried out under hypnosis is actually the same as the experience of ordinary voluntary actions, and argue that the fact that hypnosis affects both explicit verbal reports of agency and temporal binding is evidence against this alternative hypothesis. Their argument trades on the idea that hypnosis affects temporal binding via a top-down influence on motor control processes. As they put it, the underlying thought is that “movement under hypnosis represents a shift from relying on

internal action signals to relying on experimenter-delivered beliefs about action” (Lush et al., 2017, p. 662), and that this explains why temporal binding is reduced under hypnosis.

In this section, we have provided a sketch of what we take to be the broad explanatory model assumed in much of the existing literature on temporal binding, in which temporal binding is construed as a measure of an implicit or pre-reflective sense of agency. Our aim in doing so was to put on the table a plausible theoretical framework that embodies some of the claims made in the existing literature on temporal binding, which can then be compared to the alternative framework we will develop in section 4 of this paper. Some of the more specific assumptions made in our version of the model described in the current section may not be shared by all of the contributors to the relevant literature, but these differences in detail are unlikely to affect our argument in what follows. What matters to our argument is simply that any account explaining (some instances of) temporal binding as the product of a pre-reflective sense of agency must assume that instances of temporal binding occurring outside the context of voluntary action must involve the workings of an entirely separate mechanism. That is, if one wishes to hold on to the idea that temporal binding in the context of voluntary action is the product of a pre-reflective sense of agency, what we have referred to above as a two mechanisms account seems to be required once it is accepted that temporal binding can be observed in the context of causation without agency.

### **3. Temporal binding and causation**

As already mentioned, there are now a variety of studies from a number of different research groups that have demonstrated that temporal binding for two causally connected events occurs outside the context of voluntary action. Buehner (2012) showed that participants experienced a similar temporal binding of cause to effect (button press to tone) even when the cause was simply a machine depressing a button. Although the effect was not as strong as

when the button is pressed by a human participant, significant binding was still demonstrated. Temporal binding has also been observed under circumstances in which participants observe another's actions rather than carrying out an action themselves, with Poonian et al. (2015) reporting equally strong binding for observed actions and self-generated actions. Furthermore, Borhani et al. (2017) report temporal binding even when movements are entirely passive (an experimenter presses down the finger of a participant in order to initiate a tone; see also Buehner, 2015).

As mentioned previously, it has been suggested that these instances of temporal binding occur as a result of a belief in causality having a top-down influence on perception. Noting that temporal proximity can serve as a cue to causation, researchers have suggested that the psychological relation between time and causation is bi-directional (Blakey et al., 2019; Buehner, 2012; Vuorre, 2017): Not only does the timing of events influence causal beliefs; causal beliefs, in turn, have a top-down influence on the perception of the timing of events, such that events believed to stand in a cause-effect relation to each other are perceived as occurring closer together in time. (See also Ma, Chen, Kording & Wei, 2013, who draw on a similar idea to develop a Bayesian cue combination model of temporal binding.)

Our overarching question in this paper is how we should think of the relationship between the temporal binding that occurs in the context of voluntary action and the temporal binding that can also occur outside the context of voluntary action, when participants merely observe a cause-effect sequence. As we said, according to the account sketched in the previous section, these must be two separate types of binding, being brought about by two quite different mechanisms, one of which involves the purported pre-reflective sense of agency. In section 4, we will develop an alternative account, which does not appeal to the notion of a pre-reflective sense of agency, but instead construes even cases in which temporal

binding occurs in the context of voluntary actions as ones in which the mechanism by which such binding is produced is a top-down influence of causal belief on perception. By way of working towards this account, we will use this section to point out a crucial conceptual issue with the interpretation of existing findings of temporal binding in the context of voluntary action, which brings to the fore the question as to whether temporal binding, even in those cases, might not be better construed as reflecting a belief in causality. We will also describe some recent empirical evidence that makes this question particularly pressing.

### **3.1. Agency and causality: the conceptual issue**

In the existing literature that construes temporal binding as a measure of a pre-reflective sense of agency, the notion of a pre-reflective sense of agency is often associated with the idea that agents need to be able to distinguish between events they have brought about themselves and events that came about some other way. Thus, Ebert and Wegner (2010, p. 481) write that a “central problem agents face is determining which events they have caused, and which they have not”; similarly, Kühn et al. (2013, p. 1935) discuss the idea of a pre-reflective sense of agency in connection with the idea that “[g]oal-directed action requires the ability to identify the consequences of our behaviour in the external world”. Thus, the idea seems to be that of a special phenomenological feature that serves as an indicator that a particular outcome is one that has been brought about by the agent herself.

In a similar vein, Haggard and Tsakiris (2009, p. 242) also describe the sense of agency as “the experience of being in control both of one’s own actions and, through them, of events in the external world”. As this description in particular brings out, though, there are at least two separable issues that are at stake in the question as to whether one is the agent who has brought about a particular outcome. First, there is the question as to whether one’s own bodily movements and their immediate proximal consequences are one’s own doing, or

whether they are controlled by something else. In other words, the question here is whether they constitute voluntary movements. In a typical binding experiment, that question might for instance, be whether one is moving one's finger to depress a button, or whether one's finger is passively pressed onto the button by the experimenter. Secondly, there is the question as to whether one's own bodily movements and their proximal consequences are the causes of other, distal, events that may also have been brought about by some other cause. In a typical temporal binding experiment, that question might, for instance, be whether one's finger pressing on the button caused a subsequent tone.

Note that this second question, as such, is actually not a question specifically about agency. It could be rephrased as a question simply about two events – call them  $x$  (the button press) and  $y$  (the tone) – and whether or not a cause-effect relationship obtains between them, irrespective of the further question as to whether  $x$  constitutes a voluntary action. There is already some evidence that temporal binding, even in the context of voluntary action, can be influenced by factors that only bear on this question as to whether  $x$  causes  $y$  in a particular situation, such as statistical contingency (Moore et al., 2009), explicit causal beliefs about whether or not there is a causal relationship between  $x$  and  $y$  (Desantis et al. 2011), or the congruency between  $x$  and  $y$  (Ebert & Wegner, 2010). If there is a question as to “which events they have caused, and which they have not” that participants face in these studies, it is arguably simply the question as to whether  $x$  causes  $y$ , and the information participants are provided with affects temporal binding because it affects participants' causal beliefs about the relationship between  $x$  and  $y$ .

Our argument is thus that we need to draw a clear distinction, even in cases in which temporal binding occurs in the context of a voluntary action, between the claim that it is a measure of a pre-reflective sense of agency and the claim that it is a consequence of a belief that the two events that are being bound together stand in a cause-effect relationship. The

latter is simply a belief about a causal connection between two events, conceived of as two events, irrespective of the further question as to whether one of them constitutes a voluntary action.

### **3.2. Disentangling agency and causality**

In the light of these considerations, it might also be useful to return once more to Haggard et al.'s (2002) original study. The conclusions that Haggard et al. (2002) arrived at in their study were based on their belief that the only thing that varied between their conditions was the nature of the event that preceded the tone, i.e. whether it was a voluntary action, an involuntary movement, or an audible click. However, as highlighted by Buehner (2015), what also varied between the three conditions used by Haggard et al. was the presence of a plausible cause-effect structure between the two events participants had to time. It is true that the three conditions were alike in so far as the same relations of approximate temporal contiguity and contingency obtained between the stimuli in each case, but it is well-known that causal induction can also be heavily influenced by domain-specific prior beliefs, which are typically regarded as constraining the set of events that are regarded as candidate causes (Ahn, Kalish, Medin, & Gelman, 1995; Griffiths & Tenenbaum, 2009). This is particularly the case in situations in which there are few other cues to causal structure. We believe that in Haggard et al.'s (2002) study, the nature of the initial event in each condition is likely to have led participants to infer (indeed, correctly) a cause-effect structure only in one of them – the voluntary action condition in which the participants pressed a button that resulted in a tone – and a common cause structure in the other two. Participants can be expected to be familiar with the idea that a button press can cause things (even after a delay – think pressing the button on a pedestrian crossing; see also Bechlivanidis, Schlottmann & Lagnado, 2019). By contrast, if participants simply experience one of their hands twitching involuntarily and then

subsequently hear a tone, in the absence of any discernible mechanism that might connect these two events, it is much less likely that they will perceive the former to be the cause of the latter, rather than perceiving them as two separate events both brought about by the researcher running the experiment (which was indeed the case in Haggard et al.'s experimental set-up). Similarly for hearing the click of a sham TMS and then hearing a tone. As such, it may actually be the belief in a cause-effect relationship between two events that drove the observed binding effect, rather than intentional action more specifically.

That studies aiming to show effects of voluntary action on temporal binding have to avoid confounding agency and causality is now typically acknowledged. The general theoretical approach that typically informs such studies is exemplified by Borhani et al. (2017, p. 882), who still endorse the claim that temporal binding can act as a measure of a sense of agency, but restrict this claim to “a component of binding attributable to intentional action, over and above other factors” – a component which emerges when comparing temporal binding in a condition involving intentional action with the temporal binding that also occurs in a suitable control condition not involving intentional action but just the observation of two causally connected events. Thus, on this view, there are further factors, other than sense of agency, that can determine whether temporal binding occurs. This is framed, though, as an empirical rather than a theoretical challenge, i.e., it is taken to imply that experimenters using temporal binding as a measure of sense of agency need to exclude these factors.

Borhani et al. (2017) do not consider the possibility that an explanation of temporal binding on which it is construed as a measure of causal belief might actually provide an *alternative* to explanations of temporal binding that invoke the idea of a pre-reflective sense of agency. A recent study showing that this possibility needs to be taken seriously comes from Suzuki et al. (2019). They used a virtual reality set-up with tactile stimulation to match

perceptual stimuli for temporal and spatial information in a condition involving intentional action and a condition involving mere observation of a causal sequence, and found binding effects of identical magnitude across both types of condition. As they argue, it could thus be that the differences in temporal binding found in other studies featuring both an intentional action condition and a mere observation condition are not due to agency making an additional contribution to binding, but rather to other factors – e.g., the presence of additional cues in the intentional action condition that increase the precision of event timing.

We therefore think that it is at least worth considering the idea that temporal binding in general is a measure of causal belief, and in the next section, we will outline an alternative theoretical framework elaborating this idea. As Suzuki et al. (2019) argue, “[i]n-principle protection for the position that binding reflects agency has previously been provided by the fact that binding effects for non-intentional conditions have never produced a similar magnitude effect to those from intentional action conditions. [...] Our present results demonstrate that such a position is no longer defensible.” One issue one might raise about their argument, though, is how plausible it is to claim that all discrepancies in temporal binding found in other studies comparing an intentional action condition with a mere observation condition can be explained in the way they suggest. We will therefore look in particular at the question as to whether an account that construes temporal binding as a measure of causal belief has additional resources to explain such discrepancies. As we will explain, there are in fact existing considerations from the literature on causal induction that should lead us to expect that facts about the extent to which one is actively involved in bringing about an event can influence beliefs about whether or not that event is the cause of another event. If temporal binding is a measure of causal belief, this could provide an explanation as to why binding is stronger in certain conditions involving voluntary action that

turns only on factors about causal inference, rather than appealing to the idea of a pre-reflective sense of agency in the way the model sketched in section 2 does.

#### 4. Towards an alternative conceptualization of temporal binding

In Figure 2, we sketch an alternative to the theoretical framework we described in section 2, which incorporates the idea that temporal binding, even when it occurs in the context of voluntary agency, is in fact the result of a belief in causality. Again, we should stress that this sketch of an alternative explanation of temporal binding and its relation to self-attributions of agency constitutes a considerably simplified model.

Whilst this alternative model is also meant to include cases of binding occurring in the context of voluntary action, it differs from the one described in section 2 in a number of crucial respects. Two differences are particularly noteworthy. First, nothing equivalent to the

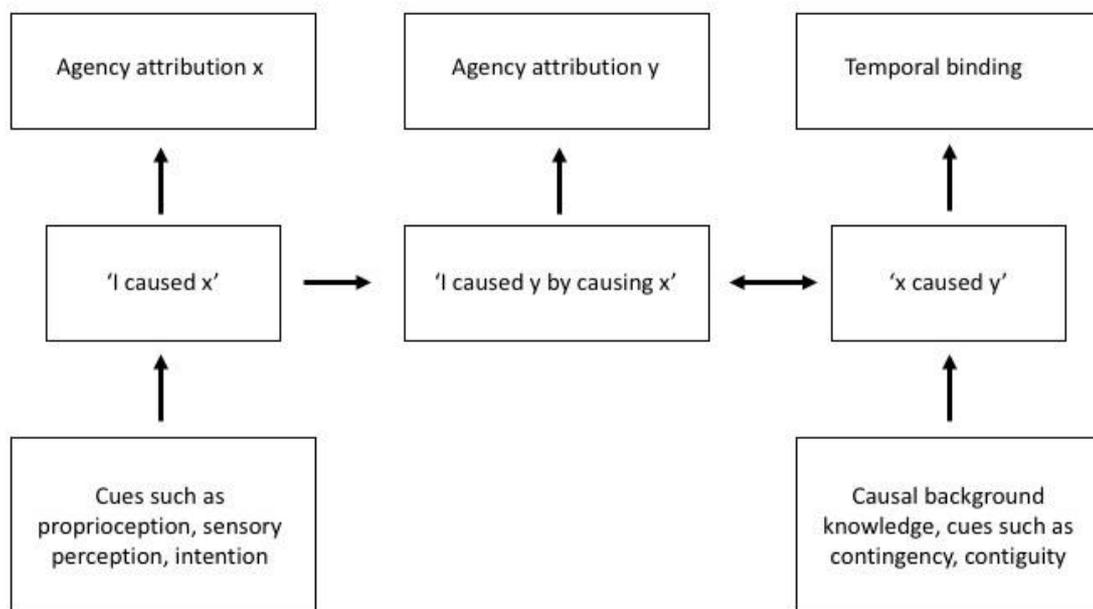


Figure 2:

Proposed alternative theoretical framework for explaining temporal binding both within and outside the context of voluntary action.

“pre-reflective sense of agency” postulated in the previous model features in this alternative model. The new model is neutral as to whether or not such a thing exists (though see Grünbaum, 2015; Paglieri, 2013); it just implies that temporal binding can be explained without appealing to this notion. Secondly, the variable assumed to underlie binding in this new model is a belief in a causal relation between the two events that are bound together, rather than anything specifically to do with agency as such (though we will shortly return to ways in which agency might nevertheless be relevant to binding on this model too).

#### **4.1. Two different types of agency judgements.**

In contrast to the previous model, this new model also explicitly distinguishes between two different types of judgements about agency, which in turn reflect two different kinds of agency beliefs. In doing so, it makes explicit the distinction we drew in the preceding section between the two separate issues that might be at stake in the question as to whether an event was due to oneself or came about in some other way. The causal belief listed on the left – “I caused  $x$ ” – is meant to correspond to a scenario in which what is at issue is the question as to whether one intentionally carried out a particular bodily movement (along with causing its immediate proximal consequences) as contrasted, for instance, with one’s body having been moved passively. (I.e., ‘ $x$ ’ here stands for the bodily movement and its immediate proximal consequences.) As an example, the question here may be whether one actively pressed a button, or whether one’s finger was moved passively to depress the button. By contrast, the causal belief listed in the middle – “I caused  $y$  by causing  $x$ ” – corresponds to a scenario in which what is at issue is the question as to whether it was one’s action that caused a distal event, rather than that event having some other cause – for example, whether by pressing the button one also caused a subsequent tone. (I.e., ‘ $y$ ’ stands for the distal event, such as a tone.)

What grounds these judgements about agency? Our model identifies two different classes of cues or background knowledge that ground people's agency judgements, which we take to be sufficient to explain people's ability to make such judgements (when they can do so, which is of course not always the case). The first are the cues underlying the judgement "I caused  $x$ ", where this refers to the judgement that one actively brought about a particular bodily movement and its immediate proximal consequences, rather than, say, one's body having been moved passively. We assume that, in making such judgements, people can rely on a number of different types of cues – cues associated with what, at the beginning of section 2, we referred to as a sense of agency in the broad sense, rather than specifically the narrower notion of a pre-reflective sense of agency. For instance, in the case of being moved passively, a conscious intention to carry out the movement will be lacking, the movement may also be unexpected and be experienced as unpredictable, and in addition one will typically be aware of sensations of resistance and pressure of a kind not encountered when intentionally carrying out the movement oneself. Conversely, in a case in which one does intentionally move one's body, one's judgement that one does so can be grounded in one's intention to do so, combined with the proprioceptive awareness of one's bodily movements and muscular activity, as well as one's sensory awareness of the immediate sensory consequences of the interactions between one's body and the environment.

This leaves the second type of agency judgement – "I caused  $y$  by causing  $x$ " – which is a judgement of agency regarding a distal event. As we noted before, such explicit judgements of agency regarding distal events ("I caused  $y$  by causing  $x$ ") quite generally need to be seen as reflecting a combination of the belief underlying the first type of agency judgement regarding one's bodily movements and their immediate proximal consequences ("I caused  $x$ ") together with the further belief that one's bodily movement and its proximal consequences caused a distal event (" $x$  caused  $y$ "). As we explained, the latter is a simply a

belief regarding a causal connection between two events, which an individual can hold also without thinking of the relevant events as having been brought about by herself. Thus, as far as what grounds that latter belief, we assume that it is formed on the basis of the usual cues appealed to in the literature on causal learning, such as information about covariation, contingency information, or existing causal background knowledge, such as knowledge about a particular causal mechanism connecting  $x$  and  $y$  or knowledge of causal generalizations (see, e.g., Lagnado, Waldmann, Hagmayer, & Sloman, 2007).

#### **4.2. Temporal binding as a measure of causal belief**

Turning now again to temporal binding, the new model thus states that such binding reflects just such a causal belief – “ $x$  caused  $y$ ” – which in the example we have been considering would be the belief that the button press caused the tone. As this type of belief is one that an individual could also have in the absence of attributing  $x$  and/or  $y$  to their own agency, our model straightforwardly incorporates findings showing that temporal binding can also occur in connection with cause-effect pairs that do not involve agency.

If our model construes temporal binding as a measure of causal belief, does that imply that binding does not measure anything of interest that could not also be measured simply by asking participants whether they believe that there is a causal connection between the two events? To some extent, this is an empirical question that is as yet unanswered because to date temporal binding studies do not typically include asking participants the relevant causal questions. However, it is not difficult to think of reasons why temporal binding, even if construed as a measure of causal belief, might serve as a more “implicit” measure that can pick up things not picked up by the “explicit” measure of simply asking people causal questions. As we will discuss further below, human causal inference is governed by a number of different principles that are typically opaque to the reasoner herself. This itself makes it

possible that people's intuitive belief about a causal relation differs from what they would say if asked for explicit judgments about the causal connection between two events. Moreover, as we will also discuss, humans have a tendency to identify one event as *the* cause of a given effect, making them more biased towards giving binary answers to causal questions (especially about singular causal relationships, which is what binding studies involve), whereas temporal binding might serve as a more sensitive continuous measure of the *degree* to which participants regard one event to be the cause of another. We will return to potential examples of temporal binding being thus sensitive to perceived differences in degree of causality below.

The idea that there are more “implicit” and more “explicit” ways in which people make causal inferences, and that the results of these can diverge from one another, is already familiar from research on causal induction. As Kuhn and Dean (2004, p. 261) argue, “individuals have available a repertory of different inference strategies or rules (reflecting different criteria for inferences of causality) from which they select variably across occasions”, potentially leading to discrepancies in their responses to different ways of probing their causal beliefs. On our proposed account, temporal binding, whilst being construed as a measure of causal belief, might constitute a measure sensitive to aspects of human causal inference that are not readily apparent from people's explicit causal judgements. In other words, what we are calling causal belief, too, might come in a more pre-reflective variety that is not easily accessible to verbal report but still expresses itself through temporal binding.

As we have said, our model can straightforwardly accommodate the finding that binding for a cause-effect pair of events can also occur in the absence of agency, because it construes binding as a measure of a belief in causality, rather than, more specifically, a pre-reflective sense of agency as assumed in the model presented in section 2. However, as we

have noted previously, temporal binding is typically stronger in the context of voluntary action, and there are a variety of manipulations affecting higher-level processes that have an influence on explicit judgments about agency, but also, at the same time, influence temporal binding. In the model presented in section 2, the influence of these higher-level processes on temporal binding was explained in terms of top-down influences on motor processes, which in turn were hypothesized to impact on the purported pre-reflective sense of agency assumed to underpin temporal binding. Since our model does not appeal to the idea of a pre-reflective sense of agency it cannot explain these effects in this way. Thus, the key question we now turn to is how our alternative model can explain findings that suggest that binding is stronger in the context of voluntary action and that it is (at least sometimes) affected by variables that also influence explicit judgements of agency.

### **4.3. Explaining the effects of agency**

We believe that there is a natural way in which our model, too, can account for the influence of higher-level processes that typically affect explicit judgements of agency, and of agency in general, on temporal binding. The element of our model that they have an influence on are the beliefs about agency (“I caused  $x$ ” and “I caused  $y$  by causing  $x$ ”). Thus, for the relevant findings to fit in with our model, these beliefs about agency must themselves be able to influence beliefs about the causal connection between two events. This is represented in our model in terms of a two-way interaction between the belief underlying agency attribution regarding the distal event (“I caused  $y$  by causing  $x$ ”) and the causal belief construed to underlie temporal binding (“ $x$  caused  $y$ ”). That is to say, whilst the causal beliefs that underlie temporal binding can feed into beliefs about agency – by allowing agency attributions to be extended to distal effects, rather than just covering bodily movements and their immediate proximal consequences – our model envisages that there is, in turn, also a way in which

beliefs about agency can influence causal beliefs. (We note here that there might perhaps also be additional reasons why the interaction between the two different first-person agency beliefs, too, should be construed as bidirectional rather than unidirectional, as depicted in our model. But we can set this question aside for present purposes.)

How exactly might beliefs about agency have an influence on the degree to which one particular event  $x$  is judged to be the cause of another particular event  $y$ ? We believe that the existing literature on causal judgement already contains the materials for spelling out ways in which this might happen. We focus on two issues in particular, which are sometimes referred to as *structure induction* and *causal selection*, respectively.

Structure induction (Lagnado, Waldmann, Hagmayer & Sloman, 2007; Meder et al. 2014, see also Griffiths & Tenenbaum 2005, 2009) refers to the individual's confidence that there is a cause-effect relationship between  $x$  and  $y$  in the first place, rather than the causal structure of the situation being one in which  $x$  does not in fact cause  $y$ . Thus, the question at issue in structure induction is the binary one as to whether or not  $x$  causes  $y$ . Yet, even in cases in which  $x$  is clearly amongst the factors involved in  $y$ 's occurrence, there may be reasons as to why it is not considered *the* cause of  $y$ . In most causal scenarios, there are a variety of factors that are involved in bringing about the effect, but people typically judge only one or a small number of them to be the cause (or causes) of the effect, whereas the rest are considered to be background conditions, or at any rate causes to a lesser degree. This is what is referred to as causal selection (Hesslow, 1988; Hilton, 2017; Lewis, 1973).

**Causal structure induction.** Taking structure induction first, how might factors that influence people's beliefs about the extent to which they are themselves responsible for an event  $x$  also have an influence on their beliefs about there being a cause-effect relationship between  $x$  and a further event  $y$ , rather than there being some other type of causal structure in place? When there is a correlation between an event  $x$  and a subsequent event  $y$ , the most

likely alternative causal structure to a cause-effect relationship between  $x$  and  $y$  is one in which  $x$  and  $y$  in fact have a common cause, rather than being directly causally related (compare our discussion of Haggard et al.'s (2002) study in the preceding section). Note that, in the context of an experiment that involves, for instance, a button press followed by a tone, the presence of such a common cause structure would imply that it is not in fact the participant that actively presses the button, but rather that their bodily movement and the depressing of the button are actually caused by something other than the participant, which also causes  $y$  via a separate causal route. Thus, conversely, the more confident the participant is that she is actively bringing about  $x$ , the more she will also judge that she can rule out the possibility that  $x$  and  $y$  in fact have a common cause (see Hagmayer, Sloman, Lagnado, & Waldmann, 2007).

This is one way of explaining the results of Borhani et al.'s (2017) study, in which they found binding of a different magnitude in a condition in which participants actively pressed a button compared to a condition in which the button push was achieved by the experimenter pressing down the participant's finger. Borhani et al. explain this difference in magnitude in terms of the idea of a pre-reflective sense of agency making a separate contribution to binding that is lacking in the condition involving just a passive movement (i.e., that there is an additive effect of two separate mechanisms). But another way of explaining the difference is that the participant's belief that the button press causes the tone is stronger in the voluntary action condition: The fact that the button press is under their control makes it easier in this condition to rule out an alternative, common cause, structure, in which it is in fact the experimenter who brings about both the button press and, separately, the subsequent tone.

Related ideas are familiar from philosophical discussions of interventionist approaches to causation (see, e.g., Woodward, 2003). Put briefly, according to

interventionism,  $x$  causes  $y$  if there is an invariant relationship between  $x$  and  $y$  that holds under a range of interventions, where for something to count as an intervention requires a specific set of formal criteria to be met. In particular, for something to count as an intervention in this sort of case, its impact on the causal system that consists of  $x$  and  $y$  must be “surgical”: ideally, it must have an influence only on  $x$ , without also itself having any influence on  $y$  other than by influencing  $x$ , and it must have an influence on  $x$  that decouples  $x$  from any other factors that could otherwise influence  $y$ . In the literature on interventionism, it is often noted that humans, in so far as they conceive of themselves as agents, think of their own actions as meeting the definition of an intervention, thus understood. In so far as they reason along the lines of interventionist principles, it is therefore to be expected that they have more confidence in their structure induction judgements in situations in which they also judge themselves to be the agents who bring about certain events (see also Lagnado & Sloman, 2002; 2004; 2006).

Thus, general principles of causal structure induction might be able to explain why some manipulations that affect participants’ judgement about the extent to which they are the cause or in control of the events they observe also affect the strength of temporal binding, even if binding itself is simply a measure of causal belief.

**Causal selection.** Turning to causal selection, we now consider how the principles governing this aspect of causal cognition, too, might play a role in determining the strength of temporal binding under particular conditions.

Principles of causal selection might, for instance, explain the results of a study by Berberian et al. (2012) examining binding in a partially automated setting. This study used a flight simulator, and participants’ task was to change a virtual aeroplane’s heading using an autopilot interface. There were four different conditions, which differed in the extent to which aspects of the change of heading were under the control of the participant or

determined by the autopilot – for instance, in one condition participants could decide on the new heading themselves, whereas in another the new heading was selected by the autopilot, and participants just had to press a button to execute the change of heading. In each condition, participants had to judge the interval between the final button press that executed the change of heading and a feedback signal indicating that the change of heading had been successfully carried out, and Berberian et al. found stronger temporal binding in those conditions in which more factors of the decision were under the participants' control as compared with conditions in which more of the factors were determined by the autopilot. In this experiment the final button press is clearly only one causal factor that determines the final result, and the different conditions are likely to have an impact on the extent to which participants will judge it to be *the* cause of that result (rather than what simply enables the autopilot to carry out the change of heading). On the model that we are proposing, this would explain the different strengths of binding Berberian et al. found in the different conditions.

Similarly, principles of causal selection might also explain the findings obtained in studies in which coercing participants to perform an action was shown to reduce temporal binding of that action to its outcome when compared to trials in which they were free to perform the same action or not (Caspar et al., 2016).

The principles people use to single out a particular cause as *the* cause of a given effect are typically studied in situations in which several different factors jointly contribute to bringing about a certain effect (as is the case in Berberian et al.'s (2012) study). However, problems of causal selection can also arise in the context of causal chains. In so far as people have a tendency to regard causal relations in causal chains as transitive (see, e.g., von Sydow, Hagmayer, & Meder, 2016), the different events in the chain compete with each other for the status of being the cause of the overall outcome. Cases of coercion are examples of causal chains, as the experimenter causes the participant to carry out a certain action by issuing the

instruction to do so, with the participant's action then causing the effect. Thus, if binding is a measure of causal belief, we might find reduced binding in such cases simply because the participant does not regard their own action as the cause of the effect, but instead traces the effect back to its root cause, the experimenter's instruction. (There is already some evidence from somewhat different studies that people have a strong tendency to trace an effect to its root cause, particularly in the context of immoral behaviour; see Sousa, Holbrook & Swiney, 2015.)

There is a sense in which all binding studies can be regarded as involving a causal chain, in so far as participants act under the experimenter's instruction, and other studies, too, might be seen as manipulating the extent to which participants will regard the causal relations in that chain as transitive, thus influencing the degree to which they see themselves as the cause of the outcome. A study by Barlas et al. (2018), for instance, examined temporal binding in an "instructed choice" and a "free choice" condition. In the instructed choice condition, participants had to press one button displayed on a screen, whereas in the free choice condition, they had a choice of four different buttons. In either condition, they were asked to respond as quickly as possible as soon as the relevant button or buttons were displayed on the screen. The difference between the instructed choice condition and the free choice condition is again likely to have affected the extent to which participants saw themselves as the cause of subsequent events rather than just a link in an unfolding chain of events. On our model, this would explain why Barlas et al. found that increasing the number of buttons participants could choose from also increased the magnitude of temporal binding.

Note that, on this analysis, it could still be said that the relative lack of binding in cases involving coercion or a lack of choice alternatives is due to the fact that the participants do not consider themselves responsible for the effect or as acting freely. Similarly, this analysis need not be seen to be in conflict with claims to the effect that coercion induces "a

cognitive operation of ‘distancing,’ or reducing the linkage between one’s own decision-making, action, and outcome” (Caspar et al., 2016, p. 590). The idea would rather be that the meaning of claims such as this can be unpacked by appealing to general principles governing causal selection in causal chains (on this, see also Deery & Nahmias, 2017), without needing to invoke the idea of a pre-reflective sense of agency.

There are, thus, ways in which an account on which temporal binding is construed as a top-down effect reflecting a belief in causality can nevertheless also accommodate findings that show that the extent to which participants are themselves actively involved in events influences temporal binding. As should be clear, though, the relevant explanations are structurally very different from ones involving the idea that temporal binding acts as a measure of a pre-reflective sense of agency, as discussed in section 2. As mentioned towards the beginning of this section, our proposed account assumes that there are numerous different factors that ground people’s agency self-ascriptions, and where the account connects temporal binding with agency it does not single out any one of them as having any privileged role. Furthermore, on our account, factors connected with agency have an impact on binding only *via* having an impact on beliefs regarding a causal connection between the two events that are bound together.

The proposed explanations, however, also provide an impetus for further empirical research. For instance, temporal binding has been shown to be affected in a wide variety of conditions such as schizophrenia (Graham-Schmidt, Martin-Iverson, Holmes, & Waters, 2016; Voss et al., 2010), Parkinson’s disease (Moore et al., 2010), autism (Sperduti, Pieron, Leboyer, & Zalla, 2014), and after administration of ketamine (Moore et al., 2011). Our account allows for the possibility that there might be different mechanisms through which these effects might occur, and thus that they might not have the same explanation across these different populations. For instance, it could be that these conditions affect the cues that

ground the belief “I caused  $x$ ” (or the interpretation of such cues), it could be that they cause impairments in the ability to use the cues grounding the belief “ $x$  caused  $y$ ”, or it could be that they have an impact on the principles governing causal selection and structure induction.

Thus, our model provides a framework for exploring which particular such factors result in reduced temporal binding in each of those populations, and in particular whether the effect is due to factors specifically connected with agency, or to more general difficulties with causal inference and/or causal perception. (See, e.g., Morris, Griffiths, Le Pelley, & Weickert, 2013; Tschacher & Kupper, 2006 for evidence that the latter are impaired in schizophrenia.)

Our account also gives rise to specific predictions that might be tested in other empirical work. In particular, we have suggested that general principles of causal structure induction and causal selection might be used to explain how factors affecting agency can also influence temporal binding. The relevant explanations would thus be undermined, for instance, if it could be demonstrated that such factors affecting agency can have an influence on temporal binding even in cases where it is implausible to think that they also have an effect on structure induction. Similarly, there are a variety of principles involved in causal selection other than those appealed to in our suggested explanations invoking influences of causal selection on binding (Hilton, 2017; Hitchcock & Knobe, 2009; Lagnado & Gerstenberg, 2017). Thus, if our hypothesis that the relevant manipulations affect temporal binding by affecting causal selection is to stand up, temporal binding should also be affected in situations in which these other principles come into play, and the hypothesis would be discredited if temporal binding and causal selection could be shown to come apart (for instance, if binding could be found under conditions in which the action is clearly just an enabling condition).

## 5. Conclusion

Since Haggard et al.'s (2002) influential work, the dominant view in the literature on temporal binding has been that temporal binding can serve as a measure of a pre-reflective sense of agency. As we have seen, proponents of this view can and typically do acknowledge that temporal binding can also occur outside the context of voluntary action. However, they need to assume that such cases involve the operation of a separate mechanism – what we have referred to as a two mechanisms account.

In this paper, we have sought to develop an alternative unitary account of temporal binding, on which temporal binding involves a top-down influence of causal belief on perception. Such an account can naturally explain cases in which temporal binding occurs outside the context of voluntary action. However, as we have also tried to show, such an account also has resources to explain why temporal binding is normally stronger in the context of voluntary action, and why it can be influenced by factors that affect the extent to which participants see themselves as being actively involved in bringing about the relevant events. Thus, the model we have presented in section 4 of this paper can serve as a unitary model for explaining temporal binding both within and outside the context of voluntary action. At this stage, this alternative model must clearly still be regarded as speculative. As we have noted, though, it can draw on existing empirical findings that show that manipulating the extent to which people see themselves as being actively involved in the events they observe does not just affect their judgements of agency, but can also affect their judgements about the causal relationships between those events themselves.

We have suggested that a strength of this model is that it provides a unitary way for explaining temporal binding in different contexts. A further strength is that it also allows temporal binding to be explained within the same general overarching framework as temporal reordering effects (Bechlivanidis & Lagnado, 2013; 2016), namely as a result of top-down

effects of causal beliefs resulting in subjective time being aligned with basic Humean assumptions about causation (see also Vuorre, 2017). That is, both temporal reordering and temporal binding can be seen as reflecting bidirectional relations between time and causation. Just as temporal cues (namely, temporal order and contiguity) can shape the acquisition of causal beliefs, so can those beliefs shape subjective time itself. Viewing temporal binding in this way does not in itself resolve outstanding issues about the underlying mechanisms (most notably, the extent to which the effect reflects the operation of basic processes involved in time perception itself, e.g., Fereday, Buehner, & Rushton, 2019; Wenke & Haggard, 2009); we note that there also parallel issues about the underpinnings of the temporal reordering effect (Bechlivanidis & Lagnado, 2016). However, positioning the temporal binding phenomenon firmly within the broader literature on time and causation provides a way of viewing it as part of a general pattern in cognition whereby existing beliefs and assumptions shape subjective experience.

## References

- Aarts, H., Bijleveld, E., Custers, R., Dogge, M., Deelder, M., Schutter, D., & van Haren, N. E. M. (2012). Positive priming and intentional binding: Eye-blink rate predicts reward information effects on the sense of agency. *Social Neuroscience*, *7*(1), 105-112.  
doi:10.1080/17470919.2011.590602
- Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition*, *14*(3), 439-458. doi:10.1016/j.concog.2004.11.001
- Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*(3), 299-352.  
doi:10.1016/0010-0277(94)00640-7
- Barlas, Z., Hockley, W. E., & Obhi, S. S. (2018). Effects of free choice and outcome valence on the sense of agency: evidence from measures of intentional binding and feelings of control. *Experimental Brain Research*, *236*(1), 129-139. doi:10.1007/s00221-017-5112-3
- Barlas, Z., & Obhi, S. S. (2013). Freedom, choice, and the sense of agency. *Frontiers in Human Neuroscience*, *7*, 514. doi:10.3389/fnhum.2013.00514
- Bechlivanidis, C., & Lagnado, D. A. (2013). Does the “Why” Tell Us the “When”? *Psychological Science*, *24*(8), 1563-1572. doi:10.1177/0956797613476046
- Bechlivanidis, C., & Lagnado, D. A. (2016). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition*, *146*, 58-66.  
doi:10.1016/j.cognition.2015.09.001
- Bechlivanidis, C., Schlottmann, A., & Lagnado, D. A. (2019). Causation without realism. *Journal of Experimental Psychology: General*, *148*(5), 785-804.  
doi:10.1037/xge0000602

- Beck, B., Di Costa, S., & Haggard, P. (2017). Having control over the external world increases the implicit sense of agency. *Cognition*, *162*, 54-60.  
doi:10.1016/j.cognition.2017.02.002
- Berberian, B., Sarrazin, J.-C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. *PloS One*, *7*(3), e34075.  
doi:10.1371/journal.pone.0034075
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*(6), 237-242. doi:10.1016/s1364-6613(02)01907-1
- Blakey, E., Tecwyn, E. C., McCormack, T., Lagnado, D. A., Hoerl, C., Lorimer, S., & Buehner, M. J. (2019). When causality shapes the experience of time: Evidence for temporal binding in young children. *Developmental Science*, *22*(3), e12769.  
doi:10.1111/desc.12769
- Borhani, K., Beck, B., & Haggard, P. (2017). Choosing, doing, and controlling: Implicit sense of agency over somatosensory events. *Psychological Science*, *28*(7), 882-893.  
doi:10.1177/0956797617697693
- Buehner, M. J. (2012). Understanding the past, predicting the future: Causation, not intentional action, is the root of temporal binding. *Psychological Science*, *23*(12), 1490-1497.
- Buehner, M. J. (2015). Awareness of voluntary and involuntary causal actions and their outcomes. *Psychology of Consciousness: Theory, Research, and Practice*, *2*(3), 237-252.
- Caspar, Emilie A., Christensen, Julia F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current Biology*, *26*(5), 585-592.  
doi:10.1016/j.cub.2015.12.067

- Chambon, V., Sidarus, N., & Haggard, P. (2014). From action intentions to action effects: how does the sense of agency come about? *Frontiers in Human Neuroscience*, 8, 320. doi:10.3389/fnhum.2014.00320
- David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition*, 17(2), 523-534. doi:10.1016/j.concog.2008.03.004
- Deery, O., & Nahmias, E. (2017). Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood. *Philosophical Studies*, 174(5), 1255-1276. doi:10.1007/s11098-016-0754-8
- Demant, J., Muhle-Karbe, P. S., Lynn, M. T., Blotenberg, I., & Brass, M. (2013). Power to the will: How exerting physical effort boosts the sense of agency. *Cognition*, 129(3), 574-578.
- Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition*, 20(4), 1211-1220. doi:10.1016/j.concog.2011.02.012
- Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PloS One*, 9(10), e110118.
- Ebert, J. P., & Wegner, D. M. (2010). Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition*, 19(1), 481-489.
- Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition*, 107(2), 693-704.
- Fereday, R., Buehner, M. J., & Rushton, S. K. (2019). The role of time perception in temporal binding: Impaired temporal resolution in causal sequences. *Cognition*, 193, 104005. doi:10.1016/j.cognition.2019.06.017

- Graham-Schmidt, K. T., Martin-Iverson, M. T., Holmes, N. P., & Waters, F. A. V. (2016). When one's sense of agency goes wrong: Absent modulation of time perception by voluntary actions and reduction of perceived length of intervals in passivity symptoms in schizophrenia. *Consciousness and Cognition*, *45*, 9-23.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334-384. doi:10.1016/j.cogpsych.2005.05.004
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661-716. doi:10.1037/a0017201
- Grünbaum, T. (2015). The feeling of agency hypothesis: a critique. *Synthese*, *192*(10), 3313-3337.
- Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences*, *9*(6), 290-295. doi:10.1016/j.tics.2005.04.012
- Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196-207. doi:10.1038/nrn.2017.14
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382-385.
- Haggard, P., & Eitam, B. (2015). Introduction. In P. Haggard & B. Eitam (Eds.), *The Sense of Agency* (pp. xi-xvi). Oxford: Oxford University Press.
- Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments, and responsibility. *Current Directions in Psychological Science*, *18*(4), 242-246.
- Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 86-100). Oxford: Oxford University Press.
- Hesslow, G. (1988). The problem of causal selection. *Contemporary science and natural explanation: Commonsense conceptions of causality*, 11-32.

- Hilton, D. (2017). Social attribution and explanation. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 645–676). Oxford: Oxford University Press.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *106*(11), 587-612.
- Kawabe, T., Roseboom, W., & Nishida, S. y. (2013). The sense of agency is action–effect causality perception based on cross-modal grouping. *Proceedings of the Royal Society B*, *280*(1763), 20130991.
- Kuhn, D., & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of cognition and development*, *5*(2), 261-288.
- Kühn, S., Brass, M., & Haggard, P. (2013). Feeling in control: Neural correlates of experience of agency. *Cortex*, *49*(7), 1935-1942.
- Kumar, D., & Srinivasan, N. (2013). Hierarchical control and sense of agency: Differential effects of control on implicit and explicit measures of agency. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*, 2790-2795.
- Lagnado, D. A., & Gerstenberg, T. (2017). Causation in legal and moral reasoning. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 565-602). Oxford: Oxford University Press.
- Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. *Proceedings of the Cognitive Science Society*, *24*, 560-565.
- Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(4), 856-876.  
doi:10.1037/0278-7393.30.4.856
- Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*(3), 451-460.  
doi:10.1037/0278-7393.32.3.451

- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. (pp. 154-172). New York, NY, US: Oxford University Press.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70(17), 556-567.
- Lush, P., Caspar, E. A., Cleeremans, A., Haggard, P., Magalhaes De Saldanha da Gama, P. A., & Dienes, Z. (2017). The power of suggestion: Posthypnotically induced changes in the temporal binding of intentional action outcomes. *Psychological Science*, 28(5), 661-669. doi:10.1177/0956797616687015
- Lynn, M. T., Muhle-Karbe, P. S., Aarts, H., & Brass, M. (2014). Priming determinist beliefs diminishes implicit (but not explicit) components of self-agency. *Frontiers in Psychology*, 5, 1483. doi:10.3389/fpsyg.2014.01483
- Ma, J., Chen, L., Kording, K., & Wei, K. (2013). Bayesian cue combination model of intentional binding. In *Proceedings of the Symposium on Translational and Computational Motor Control*. Retrieved from <https://www.seas.harvard.edu/motorlab/TCMC2013/80.pdf>
- Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, 121(3), 277-301. doi:10.1037/a0035944
- Moore, J. W. (2016). What Is the sense of agency and why does it matter? *Frontiers in Psychology*, 7, 1272. doi:10.3389/fpsyg.2016.01272
- Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: a review of cue integration approaches. *Consciousness and Cognition*, 21(1), 59-68. doi:10.1016/j.concog.2011.08.010

- Moore, J. W., Middleton, D., Haggard, P., & Fletcher, P. C. (2012). Exploring implicit and explicit aspects of sense of agency. *Consciousness and Cognition*, *21*(4), 1748-1753. doi:10.1016/j.concog.2012.10.005
- Moore, J. W., Schneider, S. A., Schwingenschuh, P., Moretto, G., Bhatia, K. P., & Haggard, P. (2010). Dopaminergic medication boosts action–effect binding in Parkinson's disease. *Neuropsychologia*, *48*(4), 1125-1132. doi:10.1016/j.neuropsychologia.2009.12.014
- Moore, J. W., Turner, D. C., Corlett, P. R., Arana, F. S., Morgan, H. L., Absalom, A. R., . . . Fletcher, P. C. (2011). Ketamine administration in healthy volunteers reproduces aberrant agency experiences associated with schizophrenia. *Cognitive Neuropsychiatry*, *16*(4), 364-381. doi:10.1080/13546805.2010.546074
- Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, *18*(4), 1056-1064.
- Moreton, J., Callan, M. J., & Hughes, G. (2017). How much does emotional valence of action outcomes affect temporal binding? *Consciousness and Cognition*, *49*, 25-34. doi:10.1016/j.concog.2016.12.008
- Morris, R., Griffiths, O., Le Pelley, M. E., & Weickert, T. W. (2013). Attention to irrelevant cues is related to positive symptoms in schizophrenia. *Schizophrenia Bulletin*, *39*(3), 575-582. doi:10.1093/schbul/sbr192
- Mylopoulos, M. I. (2012). Evaluating the case for the low-level approach to agentic awareness. *Philosophical Topics*, *40*(2), 103-127.
- Obhi, S. S., & Hall, P. (2011). Sense of agency and intentional binding in joint action. *Experimental Brain Research*, *211*(3), 655. doi:10.1007/s00221-011-2675-2
- Paglieri, F. (2013). There's nothing like being free. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), *Decomposing the will* (pp. 136-159). Oxford: Oxford University Press.

- Poonian, S. K., McFadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit agency in observed actions: evidence for N1 suppression of tones caused by self-made and observed actions. *Journal of Cognitive Neuroscience*, *27*(4), 752-764.  
doi:10.1162/jocn\_a\_00745
- Saito, N., Takahata, K., Murai, T., & Takahashi, H. (2015). Discrepancy between explicit judgement of agency and implicit feeling of agency: implications for sense of agency and its disorders. *Consciousness and Cognition*, *37*, 1-7.
- Sidarus, N., Vuorre, M., & Haggard, P. (2017). How action selection influences the sense of agency: An ERP study. *Neuroimage*, *150*, 1-13.  
doi:10.1016/j.neuroimage.2017.02.015
- Sousa, P., Holbrook, C., & Swiney, L. (2015). Moral asymmetries in judgments of agency withstand ludicrous causal deviance. *Front Psychol*, *6*, 1380.  
doi:10.3389/fpsyg.2015.01380
- Sperduti, M., Pieron, M., Leboyer, M., & Zalla, T. (2014). Altered Pre-reflective Sense of Agency in Autism Spectrum Disorders as Revealed by Reduced Intentional Binding. *Journal of Autism and Developmental Disorders*, *44*(2), 343-352.  
doi:10.1007/s10803-013-1891-y
- Suzuki, K., Lush, P., Seth, A. K., & Roseboom, W. (2019). Intentional binding without intentional action. *Psychological Science*, *30*(6), 842-853.
- 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. doi:10.1016/j.concog.2007.03.010
- Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, *149*(4), 439-446.  
doi:10.1007/s00221-003-1386-8

- Tschacher, W., & Kupper, Z. (2006). Perception of causality in schizophrenia spectrum disorder. *Schizophrenia Bulletin*, *32 Suppl 1*(Suppl 1), S106-S112.  
doi:10.1093/schbul/sbl018
- von Sydow, M., Hagemayer, Y., & Meder, B. (2016). Transitive reasoning distorts induction in causal chains. *Memory and Cognition*, *44*(3), 469-487. doi:10.3758/s13421-015-0568-5
- Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain*, *133*(10), 3104-3112. doi:10.1093/brain/awq152
- Vuorre, M. (2017). On time, causation, and the sense of agency. *Journal of Consciousness Studies*, *24*(3-4), 203-215.
- Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences*, *7*(2), 65-69.
- Wen, W., Yamashita, A., & Asama, H. (2015a). The influence of action-outcome delay and arousal on sense of agency and the intentional binding effect. *Consciousness and Cognition*, *36*, 87-95.
- Wen, W., Yamashita, A., & Asama, H. (2015b). The sense of agency during continuous action: performance is more important than action-feedback association. *PLoS One*, *10*(4), e0125226.
- Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, *115*(1), 26-38.  
doi:10.1016/j.cognition.2009.10.016
- Wenke, D., & Haggard, P. (2009). How voluntary actions modulate time perception. *Experimental Brain Research*, *196*(3), 311-318. doi:10.1007/s00221-009-1848-8

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Yoshie, M., & Haggard, P. (2017). Effects of emotional valence on sense of agency require a predictive model. *Scientific Reports*, 7(1), 8733. doi:10.1038/s41598-017-08803-3

