The illusion of (un)conscious will: Can explicit feelings of control affect low-level sense of agency?

Michael Ben Yehuda

Supervised by: Dr Robert Lockie

April 2014

The illusion of (un)conscious will: Can explicit feelings of control affect low-level sense of agency?


## ABSTRACT

Intentional binding is the perceived compression of the time interval separating an action from its effect (Action-Effect Interval (AEI)) when a sense of agency (SoA) over an event is experienced. It is positively correlated with explicit judgements of agency, although whether the relation is causal remains unclear, and is argued to occur only with intentional actions. The present study tested whether explicit knowledge of control over an effect modulates time estimates of the AEI ($H_1$), and whether binding can occur for actions with reduced intentionality (e.g., speeded reactions) if external cues suggesting self-agency are present ($H_2$). 19 participants completed a speeded reaction task in which they responded to arrows on a screen by making key-presses corresponding to the arrow's direction, after having been explicitly primed to feel either in control or not over an effect that was either compatible or incompatible with their press. $H_1$ was rejected, as no significant difference between the time estimates made in the two control-priming conditions was found, $F(1, 17) = .03$, *ns*. The null hypothesis was partially rejected for $H_2$, as a marginally significant effect of action-outcome compatibility was found, with higher estimates in incompatible than compatible trials, $F(1, 17) = 3.27$, $p = .088$. The results are discussed in terms of the way internal and external cues are integrated in the computation of SoA.



| **KEY WORDS:** | **SENSE OF AGENCY** | **INTENTIONAL BINDING** | **TIME JUDGEMENT** | **ACTION PERCEPTION** | **VOLITION** |
|---|---|---|---|---|---|

## Introduction

The sense of agency (SoA) is the experience of controlling one's own actions, and through them certain events in the environment. For example, if the lights turn on upon entering a dark room, a feeling of self-agency arises over the event if immediately prior to its occurrence one had flipped the light switch with the intention of lighting up the room. Traditionally, the literature has analysed the impact of both retrospective and prospective influences on SoA (Haggard & Chambon, 2012), the former affecting authorship after an action has caused its consequence, and the latter before an action has been performed. Two moments in the production of an action seem to be crucial for the computation of agency: the moment immediately after an effect has occurred, during which an agent compares the intended sensory outcome of a movement with the actual consequences this has caused (postdictive), and the process of selecting an action through which to implement a goal (predictive).

According to one influential model of the computation of SoA – Blakemore, Frith, and Wolpert's *comparator model* (1999; Blakemore, Wolpert, & Frith, 1998; Frith, Blakemore, & Wolpert, 2000a) – the feeling of authorship is an inferred construct that is generated retrospectively once a certain effect has been caused in the environment. Specifically, they propose that an inverse model guides movements by internally representing a desired state and computing a set of motor commands that are likely to lead to the wanted outcome given the current circumstances. At the same time, a forward model generates an efference copy containing the predicted sensory consequences of a given action, and thus adjusts the sequence of movements programmed by the inverse model based on a comparison between anticipated and current states. If the final effect achieved by the action is congruent with the predicted one, then the individual infers agency over the outcome; if, instead, there is a discrepancy, an external causal attribution is made.

Wegner and Wheatley (1999) provided support for the comparator model by measuring perceived intentionality after having primed participants with the consequences that their actions would cause before these were executed. Specifically, participants were supposed to rate how much they intended to perform an action that was generated either by them or by a confederate. On trials in which the confederate performed the action, participants were primed either with its effect or with a different stimulus 30 seconds (s) before, 5 s before, or 1 s before the action occurred. Even though intentionality was null on these trials, as actions were externally caused, participants gave higher ratings for trials in which prime and outcome were congruent as opposed to incongruent. That is, activating the representation of an outcome increased judgements of self-control when the comparison between the actual outcome and the represented stimulus produced a congruent result, possibly replicating real-life situations in which an agent has a specific expectation as to the consequences of his actions. However, evidence for a prospective SoA suggest that the comparator model may only grasp one facet of the computation of authorship, and thus not be sufficient to account for the entire experience of agency (Synofzik, Vosgerau, & Newen, 2008).

The process of selecting an action has been found to affect SoA in a number of studies (e.g., Chambon & Haggard, 2012; Wenke, Fleming, & Haggard, 2010). Wenke and her colleagues manipulated fluency of action selection by requiring participants to respond to right- or left-pointing arrows on a screen by making right or left key presses accordingly. Immediately before the appearance of the target arrows, participants were unknowingly primed with subliminal arrows pointing either compatibly (i.e., in the same direction) or incompatibly (i.e., in the opposite direction) with the target arrows. On some trials, the target arrow pointed in both directions, in which case participants could freely choose which key to press. After making their response, one of a number of possible colour patches appeared on the screen, and participants were then asked to rate how much control they felt they had had over the appearance of that specific colour. Crucially, subjects had the same degree of control (none) over the appearance of each of the colours, as these varied based on the compatibility between primes and targets, and on the hand used to make the press. However, results showed that not only were participants slower and more prone to making errors in incompatibly primed actions as well as more likely to follow the prime on the free-choice trials, they also felt significantly greater control over the colours that followed compatible trials as compared to colours that were preceded by incompatible prime-target pairings. That is, when the action selection process was fluent, participants felt greater agency over the effects that followed than when contrasting cues were presented prior to the performance of the press.

It is possible that by modulating fluency of action selection, subliminal motor priming captures experimentally the experience of generating either well-trained actions that are automatic and require little planning (compatible priming), or novel actions performed with hesitation and that require preparation (incompatible priming) (Haggard & Chambon, 2012). Research on the neural basis of the SoA shows that one cortical area in particular, the angular gyrus (AG), is activated to different degrees depending on whether participants are primed compatibly or incompatibly prior to making an action. By performing a functional magnetic resonance imaging (fMRI) scan of participants while they completed a motor priming task analogous to the one used by Wenke et al. (2010), Chambon, Wenke, Fleming, Prinz, and Haggard (2013) found that as SoA decreased in prime-target incompatible trials, AG activation increased when compared to a baseline resting condition. On the other hand, no significant change in activity from the baseline was found in

AG during compatible trials, consistent with the hypothesis that SoA may be a default experience of action production with AG monitoring infractions of this working mode (Frith, Blakemore, & Wolpert, 2000b). In addition, Chambon and his colleagues found that AG activation was inversely related to activity in the dorsolateral prefrontal cortex (DLPFC), which was significantly higher than in the resting condition only during compatible trials. Interestingly, previous research found that DLPFC activation is involved in the generation of willed action (e.g., Hyder et al., 1997), suggesting that compatible motor priming may engage some of the same circuits responsible for the production of willed actions, thus replicating partially situations of true agency.

Therefore, it seems likely that agency is computed from an integration of both prospective and retrospective factors, which acquire different weightings in the computational process depending on the circumstances in which a movement is generated. Synofzik et al. (2008) propose a two-step account of agency, in which a distinction is made between an implicit, low-level, perceptual feeling of agency (FoA) and an explicit, higher-order, conceptual judgement of agency (JoA). The former represents the first step in the computational process, and it serves the purpose of classifying an action as self-caused or not self-caused based on certain internal cues (e.g., congruence between predictions and sensory outcomes of an action, as in the comparator model; fluency of action selection). However, when the FoA computation does not allow for an action to be classified as self-caused (e.g., in the case of a mismatch between proprioceptive cues, motor predictions, and sensory feedback), a more explicit JoA occurs in which explicit beliefs and external cues are incorporated, and agency is explicitly attributed either to the self or to another agent.

Moore, Wegner, and Haggard (2009) found evidence for such a Bayesian-integration model of SoA by using an effect-priming experimental paradigm. Participants were supraliminally cued with the consequence that a movement, either voluntary or involuntary (caused by the experimenter), would cause. In some cases the cue matched the actual outcome, in others it did not. The results showed a significant two-way interaction between voluntariness of movement and prime-outcome compatibility, so that matching prime-outcome pairings modulated agency on involuntary movement trials more strongly than on voluntary movement trials. This interaction supports the hypothesis that different types of agency cues are weighted more or less heavily depending on the way an action is performed. When intrinsic motoric cues to self-agency are present, as when a voluntary action is performed generating an efference copy that provides highly reliable temporal information (Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005), then external cues are less valued. Conversely, when such intrinsic cues are absent due to an action being externally generated, extrinsic cues gain more weight, thus modulating feelings of authorship to a greater extent. Consistent with Wegner and Wheatley's (1999) results, this means that extrinsic cues may override intrinsic ones, producing a feeling of ownership over an action even when the intention to act is actually missing.

One factor that affects how both prospective and retrospective influences on SoA are weighted is temporal contiguity between different cues. Specifically, the duration of the temporal delay separating the presentation of an outcome prime and the generation of a subsequent action is inversely correlated with the agency one feels over a subsequent effect (Moore et al, 2009; Wegner & Wheatley, 1999). For example, outcome primes presented 1 s before a movement is produced increase agency over an outcome more than primes presented 30 s before (Wegner & Wheatley, 1999). A similar inverse relationship exists between the temporal duration of the action-effect interval (AEI) and the perceived causal relationship between an action and its outcome (Shanks, Pearson & Dickinson, 1989). The effect is also present both when movements are primed (e.g., subliminal arrows primes) (e.g., Chambon et al., 2013; Wenke et al., 2010) and when effects are cued prior to action execution (e.g., Moore et al., 2009). Interestingly, the opposite also holds true, as events over which an individual feels agency are perceived as occurring closer in time to the actions that generated them than outcomes over which one feels no SoA (Haggard, Clark, & Kalogeras, 2002). This effect is known as intentional binding, and is commonly used in the literature as an implicit marker of SoA.

In their seminal article Haggard et al. (2002) had participants estimate the time at which they either made a key press or heard a tone using a Libet clock (Libet, Gleason, Wright, & Pearl, 1983), a task that generated fairly accurate results when each event took place alone. However, when in the operant condition key presses were followed by the tone in the same trial and participants were required to estimate the time at which each event had occurred, time estimation for the occurrence of the action was characterised by a significant positive bias of +15 milliseconds (ms) and time estimation for the occurrence of the tone showed a significant negative shift of -46 ms. Thus, the 250 ms interval objectively separating the two events was perceived as lasting only 189 ms. Crucially, when key presses were induced by transcranial magnetic stimulation (TMS), thus producing unintentional finger movements, the opposite pattern was found, as the same time interval was perceived as lasting 308 ms (Temporal estimate biases: -27 for action and +31 for effect), highlighting how the presence or the lack of intentionality affects time perception.

The validity of intentional binding as an implicit measure of SoA is further supported by research analysing the correlation between subjective shifts of time perception and explicit judgements of agency. In a series of two experiments, Ebert and Wegner (2010) required subjects to either push or pull a joystick as everyday objects appeared on a screen. Half of the time when subjects chose to pull the joystick the object seemed to get closer (compatible trials), and the other half of the time it seemed to get farther away (incompatible trials), the same pattern occurring also when participants decided to push. An AEI of either 100, 400, or 700 ms was included between the beginning of the movement and the onset of the object's motion. Subjects were then asked to estimate how long the AEI had lasted and how much control they felt they had had over the motion of the object. A significant negative correlation was found between intentional binding and explicit judgements of agency, as compatible action-effect pairings produced proportionally shorter temporal estimates and greater SoA ratings when each of the three AEIs was presented. Importantly, in their second experiment, in which the two authorship measures were presented in independent blocks (i.e., one block only temporal estimates and one block only explicit judgements of agency), a dissociation was found, as in this case the correlation was not significant due to the fact that action-effect compatibility modulated explicit judgements of agency more strongly than intentional binding. Thus, it has been hypothesised that binding measures the low-level FoA hypothesised in Synofzik et al.'s (2008) model, while explicit judgements measure the higher-order attribution of agency, namely the JoA in the multifactorial model (Ebert & Wegner, 2010; Moore et al., 2009).

The present study examines how intentional binding and explicit judgements of agency are related, and particularly whether the higher-order, explicit experience of agency (the JoA in Synofzik et al.'s (2008) two-step model) can affect the lower-level SoA measured as the degree of perceived temporal binding between an action and its effect. Specifically, by priming participants to feel either explicitly in control or not of an event in two separate blocks, the effects of higher-order attribution of agency on the time estimates participants made as to the duration of the AEI were measured. If a top-down influence from higher-order SoA to low-level feelings of authorship is possible, then participants should have experienced the AEI on the trials in which they were primed to feel in control of the effect as being shorter than in trials in which they were primed to believe they lacked control. Such an effect would also mean that the second step in Synofzik et al.'s model may not necessarily follow the first step chronologically, and that the order with which they are computed varies depending on specific circumstances.

**$H_1$**: There will be a significant difference between participants' time estimates of the AEI in trials in which they are primed to feel in control of the effect of their actions and trials in which they are primed to feel like the effect is random.

As in Ebert and Wegner's (2010) study, a more naturalistic design was chosen in the present experiment than the pressing of a key and the appearing of a colour patch that has often been used in the literature. In this case participants were required to make a right or a left speeded key press in response to right- and left-pointing arrows, which resulted in a grey colour patch moving either right or leftwards, half the time compatibly with the action and half the time incompatibly. To the author's knowledge, this is the first study to measure intentional binding in a speeded response task, as most experimental paradigms using binding (e.g., Haggard et al., 2002; Moore et al., 2009) give participants the choice of *when* to perform a given action, but not *whether* they want to perform it, *what* the action will be, and *how* it will be performed, thus allowing for a degree of intentionality in the action produced. That is because binding is believed to occur only for intentional actions where an individual has internal control over at least one aspect of the action being performed (P. Haggard, personal communication, March 16, 2014). In the present experiment participants had no freedom of choice, as also the *when* dimension was controlled by the inserting of a time constraint on the production of the key presses, while action-effect compatibility was used as an external cue suggesting possible self-agency. If an intentional binding effect were found nevertheless, that could mean that the perceived compression of the AEI can occur also in the absence of any intentionality in the action execution process, provided there are alternative external agency cues, such as the compatibility between action and effect. The use of speeded reactions was also functional to the manipulation used to test the first experimental hypothesis, as it allowed to measure the top-down effects of explicit knowledge of control without the interference (or at least with reduced interference) of uncontrolled internal agency cues that might have affected the results.

**$H_2$:** There will be a difference between time estimates in compatible and incompatible trials.

## Method

### Design

The experiment followed a 3 × 2 × 2 within-subjects design with action-effect interval (AEI) (100, 400, and 700 milliseconds (ms)), compatibility (compatible vs. incompatible), and explicit control (high-control vs. no-control) as independent variables (IV). The first two IVs varied between individual trials (i.e., different conditions of the IVs were presented in different trials), whereas the latter varied between blocks (i.e., one block of trials was wholly high-control whereas the other was wholly no-control). Counterbalancing was used in order to avoid possible confounding effects of the order in which high-control and no-control blocks were presented so that half the participants did the high-control block first and the no-control block second, and vice versa. The dependent variable was subjects' time estimates of the action-effect interval in milliseconds.

### Participants

25 participants, mostly undergraduates of the University of West London, were recruited through a convenience sampling to take part in the experiment. Data from six participants were excluded from analysis due to erratic performance on the time estimation task. Specifically, either the correlation between their time interval judgements and the actual time intervals in a series of practice trials was not significant ($n = 3$) (Ebert & Wegner, 2010) or the standard deviation of their time estimation in the experimental task was greater than 300 ms in one or more conditions ($n = 3$) (Moore et al., 2009). Thus, the final sample was made up of 19 participants (7 males; 12 females) aged on average 25.32 years (± 9.99 years). The experiment was conducted with the approval of the ethics committee of the Department of Psychology of the University of West London. All subjects were informed of the ethical guidelines and gave written consent to take part in the study prior to the beginning of the experiment.

### Materials

The experiment was designed and run using software SuperLab 5 for Mac. Testing was conducted on a MacBook Pro with a 13-inch LED screen set at 1280 × 800 resolution and a 60-Hz refresh rate, and with a 2.4 GHz Intel Core i5 processor.

The effects of participants' actions in the experimental task were two 0.75 seconds long video animations of a grey colour patch moving either rightward or leftward, created in Adobe After Effects CS5. In both videos the colour patch started its motion from the centre of the screen and moved at a constant speed either to the right or to the left of the screen until disappearing. The remainder of the stimuli used in the experiment (i.e., right- and left-pointing arrows used as cues in both the training and the experimental tasks) were designed using Microsoft PowerPoint.

**Procedure**

Participants were seated at a desk in a quiet room at a distance of approximately 60 cm from the computer monitor on which the experiment was conducted. Before beginning the experimental task subjects completed a training task aimed at familiarising them with the concept of milliseconds and teaching them how to discriminate confidently between different lengths of time between 0 and 900 ms. (This task replicated a practice task previously used by Ebert and Wegner in 2010.)

*Practice Task* – Participants were instructed to make untimed right or left key-presses in response to black right- or left-pointing arrows appearing over a white background in the centre of the screen. Half the time the arrows pointed rightward and the other half leftward. The order of presentation of right and left arrows was random. After participants pressed the appropriate key, the screen went black for a time between 0 and 900 ms (including all and only multiples of 100 ms; i.e., 100, 200, … , 900 ms), following which a white flash appeared at the centre of the screen for 34 ms. Participants were then asked to estimate how long the screen had gone black after they had made their press and before the appearance of the flash, by using a 10-point scale with each point representing an increment of 100 ms starting from 0 ms. After making their estimate, participants received feedback on the actual time interval, before moving on to the next trial. Each subject completed a block made up of a total of twenty practice trials with each time interval being presented twice in a random order.

Besides familiarising participants' with a time discrimination task using intervals of milliseconds, it was hoped that by using ten different time durations the practice trials would increase the chances that participants believed these same intervals were also used later in the experimental task (whereas in fact only three time intervals were used in that part of the study) (Ebert & Wegner, 2010). Furthermore, it allowed monitoring cases in which participants had particular difficulty with the task, informing potential decisions to exclude subjects from the experiment.

*Experimental Task* – After completion of the practice, participants began the experimental task, which was divided into two separate blocks of sixty trials each. (The trial structure is shown in Figure 1.) Participants were instructed to make speeded responses to right- or left- pointing arrows that appeared over a fixation point at the centre of the screen by pressing a right or a left key accordingly. The arrows pointed rightward 50% of the time, and leftward 50% of the time. The order in which right- and left- pointing arrows were presented was random. Target arrows remained on the screen for 500 ms during which participants had to make the correct response. If this time limit was exceeded or the response made was incorrect, a red X appeared on the screen and the trial was repeated. A 500 ms response window was chosen so as to prevent participants from consciously controlling or monitoring their reaction time (RT) during the trials, since RT was used as a manipulation in the high-control block (see below). When participants made a correct response, the arrow disappeared and the screen remained blank for a time interval of either 100, 400 or 700 ms, following which a grey colour patch appeared in the centre of the screen and moved either to the right or to the left. The three AEI durations were used as they represent standard intervals in the binding literature (e.g., Moore et al., 2009; Ebert & Wegner, 2010).

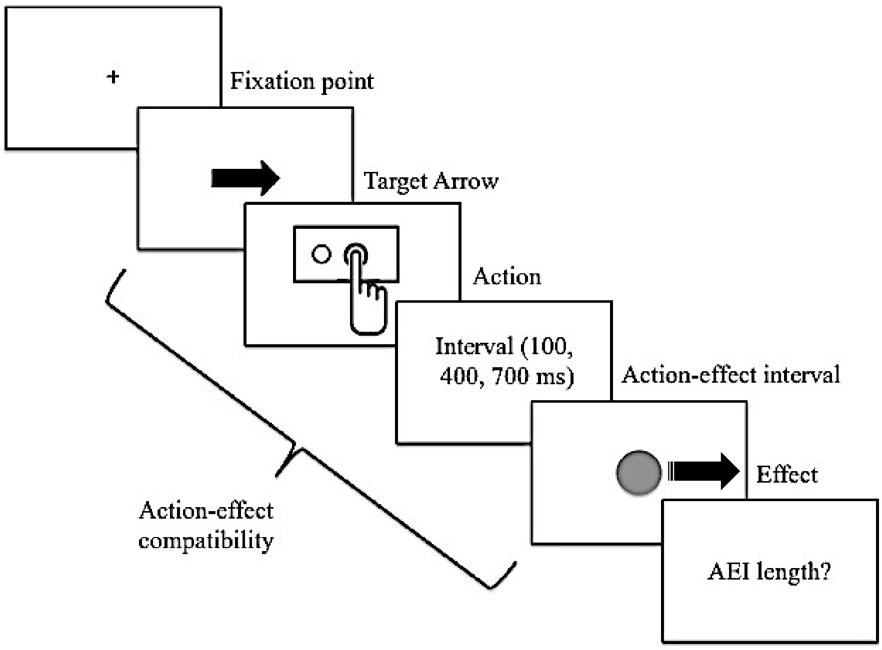


**Figure 1:** Experimental task trial structure: Participants were instructed to respond to right- or left-pointing arrows appearing on a screen by making right or left key presses accordingly. After a grey colour patch moved either right or left following an interval of 100, 400, or 700 ms, participants were required to estimate the duration of the temporal delay separating press from the onset of the movement.

Each interval appeared in one third of the trials for a total of twenty presentations each. In half the trials within each AEI condition the colour patch moved compatibly with the participant's action (i.e., it moved right when the participant pressed the right key and vice versa) and in the other half incompatibly (i.e. it moved left when the participant pressed the right key and vice versa). At the end of the effect, participants were asked to estimate how much time had passed from the moment they had made their response to the moment the colour patch started its motion, using a 10-point scale with each point representing an increment of 100 ms starting from 0 to 900 ms. Unlike in the practice trials, participants did not receive feedback on their performance.

The same trial structure was used in both blocks that participants completed, with the difference that in the high-control block participants were led to believe that they could determine the direction in which the colour patch moved, whereas in the no-control condition they were made to believe that the direction of the effect was unrelated to their action and thus completely random. Specifically, before the beginning of the high-control block participants were instructed that compatibility between their action and the effect was dependent on the speed of their actions, and that the quicker they made their response, the greater the chances that the effect would be compatible with it. Participants were further instructed that their goal for that block was to achieve as many action-effect compatible trials as possible. Conversely, before the beginning of the no-control trials, participants were instructed that the direction of the effect was completely random. Crucially, action-effect compatibility was never controlled by the participants' action nor was it random, as there was a 50-50 split for action-effect compatibility in each block of trials. The decision to use reaction time to manipulate feelings of control was given by the fact that it was found not to influence the explicit experience of agency in speeded response tasks (Chambon & Haggard, 2012).

## Results

### Preliminary Analysis

In order to measure how accurate participants were in estimating short time intervals, the data from the practice trials were analysed using Pearson's *r*, since an interval scale was used for the task. The results revealed that participants were highly accurate in their judgements, as a strong positive correlation was found between estimated temporal durations and actual temporal durations (*M* Pearson's $r = .76$). For three participants the correlation was either not significant or negative; data from these three subjects were excluded from further analysis, as this was taken as evidence of particular difficulty with estimating short time intervals.

### Binding

The data from the experimental task were analysed by obtaining the average scores of participants' responses in each of the different AEI (Action-Effect Interval), compatibility, and control combinations, resulting in twelve mean scores per participant (Figure 2). In addition, the order in which the blocks were presented was included in the analysis to check for order effects. Thus, a 2 × 2 × 2 × 3 mixed factorial ANOVA was performed on participants' time estimates, with Order of block presentation (High-control first vs. No-control first) as a between subjects factor, and Control (High- vs. No-control), Compatibility (Compatible vs. Incompatible), and AEI (100 vs. 400 vs. 700 ms) as within subjects factors. Table 1 provides the means and the standard deviations of participants' average responses across the different control × compatibility levels.

**Table 1**
**Mean time estimates in ms (SD across subjects)**

| | **High Control** | **No Control** | **Total** |
|---|---|---|---|
| **Compatibility** | Mean (*SD*) time interval estimate (ms) | | |
| **Compatible** | 376.96 (149.33) | 390.74 (151.97) | 383.85 (150.14) |
| **Incompatible** | 410.53 (172.61) | 401.54 (156.67) | 406.03 (164.16) |
| **Total** | 393.74 (161.55) | 396.14 (153.75) | |

An inspection of the means shows that overall there was only a small difference between time estimates across the two control levels, with time intervals in the no-control block being perceived as slightly longer than trials in the high-control block. Although the difference in the estimates was minor, this trend is consistent with the hypothesis being tested, according to which explicit judgements of agency can affect low-level temporal binding. However, this was true only for compatible trials, since AEIs in incompatible trials, instead, were perceived as being shorter in the no-control block than in the high-control block. There is also a difference in time estimates between trials with compatible and incompatible action-effect pairings. Specifically, on average AEIs in compatible trials were judged as being shorter than AEIs in incompatible trials in both the high- and the no-control block. Finally, as Figure 2 shows, time estimates increased as actual intervals did, with 100 ms AEIs being judged as shortest, 400 ms AEIs being judged as intermediate, and 700 ms AEIs being judged as longest across all control × compatibility conditions.

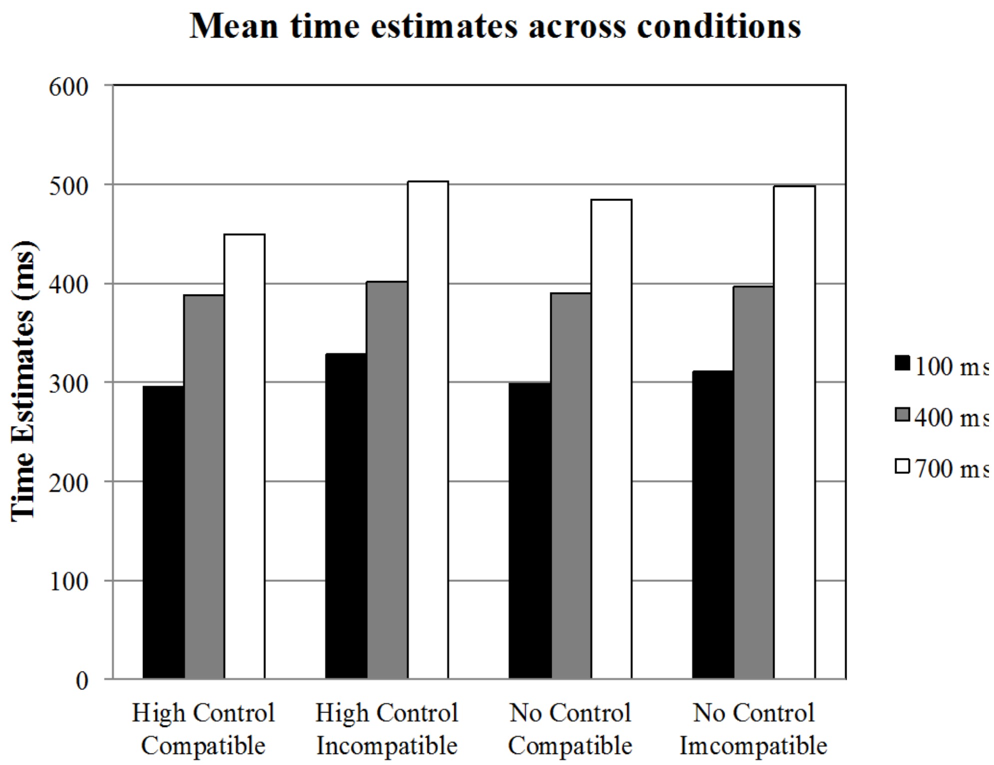


**Figure 2:** Mean time estimates across different control × compatibility × AEI conditions.

The ANOVA revealed that the main effect of the order of block presentation was not significant, $F(1, 17) = 1.16$, *ns*, meaning that time estimates did not vary significantly depending on whether the high- or the no-control block was tested first. The main effect of control was also not significant, $F(1, 17) = .03$, *ns*, indicating that explicit beliefs of control over outcomes did not significantly affect time estimates of the intervals. The main effect of compatibility between action and effect was marginally significant, $F(1, 17) = 3.27$, $p = .088$, with compatible trials yielding slightly lower time estimates than incompatible trials. Finally, the main effect of AEI was highly significant, $F(2, 34) = 24.78$, p < .001. Bonferroni corrected t-tests were used as a post-hoc test to asses which differences across estimates of the three AEIs were significant. It was found that average estimates of 100 ms intervals (*M* = 308.04 ms) were significantly lower than those of both 400 ms (*M* = 393.48 ms), *t*(17) = 6.54, p < .01, and 700 ms intervals (*M* = 483.30 ms), *t*(17) = 9.40, p < .001. In addition, 400 ms intervals were judged as being significantly shorter than 700 ms intervals, *t*(17) = 7.74, p < .001.

The order of block presentation × control interaction was not significant, *F*(1, 17) = 1.03, *ns*, meaning that the order in which the two blocks were presented did not significantly affect time estimates across the two control conditions in a different way. In addition, the order of block presentation did not interact significantly either with compatibility, *F*(1, 17) = 2.34, *ns*, or with AEI, *F*(1, 17) = .20, *ns*. The control × compatibility interaction was also not significant, *F*(1, 17) = 1.48, *ns*, suggesting that time estimates in the different compatibility conditions were not significantly affected whether the block in which they were performed was high- or no-control. Control failed to interact also with AEI, *F*(1, 17) = .38, *ns*, which means that the level of control did not have different effects on the estimates of any of the three AEIs. The same applies to compatibility, as the compatibility × AEI interaction was also not significant, *F*(1, 17) = 1.44, *ns*. Finally, none of the three- and four-way interactions was significant.

### Precision

An additional analysis was conducted on the standard deviations of participants' time estimates in each of the control × compatibility × AEI conditions, as variance within the data was considered as a measure of temporal judgement precision. It is important to note that this second analysis and any significant results that may arise should be interpreted with caution, since the analysis was run merely as a tool for exploring the data at greater depth following the interpretation of the results obtained from the ANOVA conducted on the means. As for the first test, also in this case a total of twelve scores were collected per subject, one from each of the different conditions of

**Table 2**
**Mean standard deviations in ms (SD across subjects)**

| | **High Control** | **No Control** | **Total** |
|---|---|---|---|
| **Compatibility** | Mean (*SD*) standard deviation (ms) | | |
| **Compatible** | 146.58 (54.92) | 147.28 (51.27) | 146.93 (52.89) |
| **Incompatible** | 144.36 (54.94) | 162.91 (58.59) | 153.63 (57.31) |
| **Total** | 145.47 (54.70) | 155.10 (55.37) | |

the within-subjects independent variables. These were analysed with a 2 × 2 × 2 × 3 mixed-factorial ANOVA, with both the between and the within subjects factors, as well as the levels of each, being the same as the ones of the analysis conducted on the means. Table 2 provides the average standard deviation across the different control × compatibility conditions.

On average participants were more precise in the high-control than in the no-control block and in compatible than in incompatible trials. The difference in precision in the two control conditions was more accentuated in incompatible trials, and the greater precision in compatible than in incompatible trials was found only in the no-control block. The ANOVA revealed that the main effect of control was marginally significant, $F(1, 17) = 3.09$, $p = .097$, with more precise estimates being made in the high-control block, while the main effect of compatibility was not significant, $F(1, 17) = 2.11$, *ns*. The main effect of AEI was highly significant, $F(2, 17) = 11.98$, $p < .001$. Bonferroni corrected t-tests indicated that precision in 100 ms estimates ($M = 143.14$ ms) was not significantly different than in 400 ms estimates ($M = 134.41$ ms), $t(17) = 1.41$, *ns*. 700 ms estimates ($M = 173.30$ ms) were significantly less precise than both 100 ms, $t(17) = 4.13$, $p < .05$, and 400 ms estimates, $t(17) = 5.78$, $p < .001$ (see Figure 3). Finally, the main effect of the order of block presentation was not significant, $F(1, 17) = .65$, *ns*, meaning that there were no order effects on participants' precision at estimating time intervals.

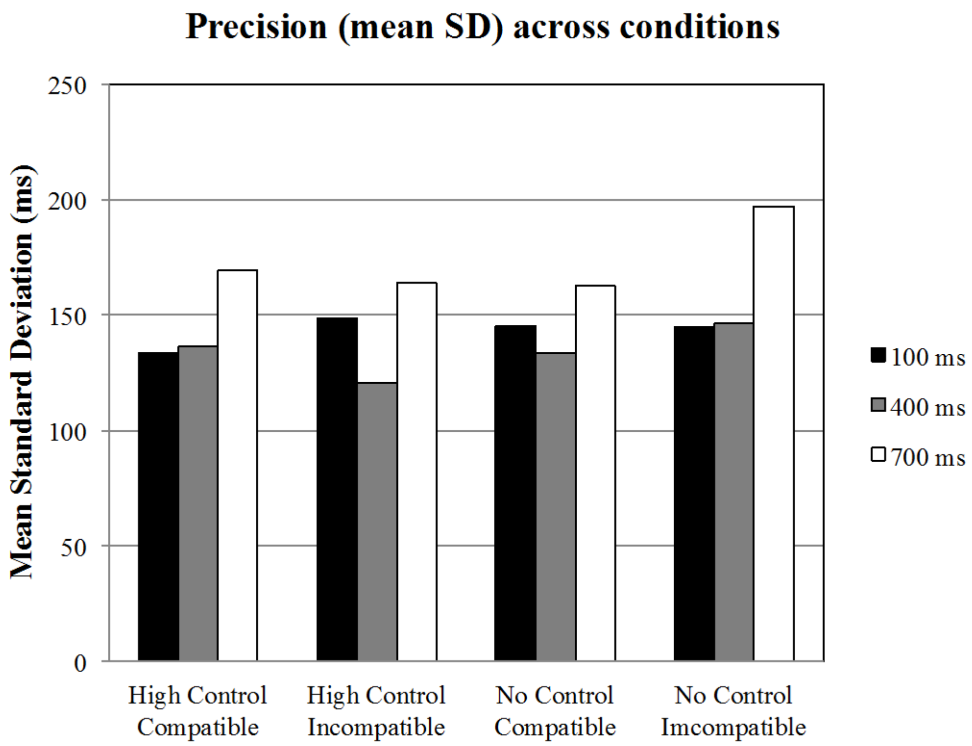


**Figure 3:** Mean precision across different control × compatibility × AEI conditions.

In addition, there were a number of significant and marginally significant two-way interactions. The order of block presentation × control interaction was significant, $F(1, 17) = 5.10$, $p < .05$, meaning that precision across the high- and no-control blocks varied depending on which block was performed first. Simple effects analysis revealed that participants were significantly less precise in the no-control than in the high-control block when the former was performed first, $F(1, 17) = 7.65$, $p < .05$, while there was no significant difference when it was performed second, $F(1, 17) = .13$, *ns*. The control × compatibility interaction was marginally significant, $F(1, 17) = 3.92$, $p = .064$, indicating that the control conditions affected precision across the two compatibility levels differently. Simple effects analysis showed that the no-control block produced significantly lower precision in incompatible than in compatible trials, $F(1, 17) = 4.88$, $p < .05$, whereas there was no significant difference in the high-control block, $F(1, 17) = .02$, *ns*.

Finally there were two marginally significant three-way interactions. The first one was the order of block presentation × control × compatibility interaction, $F(1, 17) = 3.81$, $p = .068$. Simple effects analysis revealed that participants were significantly less precise in incompatible trials within the no-control block when this was performed before the high-control block, $F(1, 17) = 10.64$, $p <.05$, but not when it was performed second, $F(1, 17) = .05$, *ns*. The second marginally significant interaction was the control × compatibility × AEI interaction, $F(2,16) = 3.09$, $p = .073$. Specifically, simple effects analysis indicated that incompatible trials caused lower precision in the no-control block with 400 ($F(1, 17) = 5.93$, $p < .05$) and 700 ms intervals ($F(1, 17) = 4.76$, $p < .05$), but not with 100 ms intervals ($F(1, 17) = .02$, *ns*).

**Reaction Time**

A final set of analyses was run on participants' reaction times (RT) across the different conditions to ascertain that there was no relationship between their RTs and their time estimates. This analysis was particularly relevant for the high-control blocks, where participants were instructed that the faster they responded to the target arrows on the screen, the more likely the effect would follow the direction of their press (i.e., the greater the chance that they would control the direction of the effect). Therefore, it was important to establish that participants were not actually relying on monitoring their RT in order to inform their time estimates, as that would have interfered with the manipulation that was used to attempt to affect their feelings of explicit control. Averages of each participant's RTs in the high- and no-control blocks were computed and compared, revealing that participants were significantly faster in responding in the high-control blocks ($M = 386.23$ s) than the no-control blocks ($M = 397.95$ s), $t(18) = 2.30$, $p < .05$, perhaps because the importance of short RTs was emphasised in the former block but not in the latter. In addition, the correlation between participants' RTs and their time estimates was computed for each of the control × compatibility × AEI conditions, generating twelve correlation coefficients per subject. Then, the coefficients of the different compatibility × AEI conditions were collapsed for analysis. The mean Pearson's *r* coefficient was .06 in both high- and no-control blocks, indicating that RTs were not related with time estimates, and thus that participants did not rely on their RT in order to produce their time estimates.

## Discussion

The present study tested the hypothesis that explicit feelings of agency over an outcome can affect the implicit sense of agency (SoA) measured by intentional binding. If this kind of top-down control of explicit over implicit authorship were possible, then participants should have felt that the action-effect interval (AEI) separating actions and effects in the block in which they were primed to feel in control of the outcome of their actions was shorter than in the block in which they were primed to feel like the outcome was due to chance. This first experimental hypothesis was rejected, as there was no significant difference in the time estimates participants made in high- and no-control blocks, suggesting that manipulating explicit feelings of agency as was attempted in this study does not affect the implicit experience of authorship captured by intentional binding. The study also tested the hypothesis that a main effect of compatibility between action and effect on time estimates would be present despite the use of a speeded response task. In this case the null hypothesis was partially rejected, as a marginally significant binding effect modulated by action-effect compatibility was found, with compatible pairings causing slightly lower time estimates than incompatible pairings. This shows how binding may occur even when speeded reactions are used instead of intentional actions if appropriate external cues suggesting potential self-agency are present.

There are two main explanations available for the lack of an effect of explicit SoA over time estimates that led to the rejection of the first experimental hypothesis: either (1) a top-down influence of explicit agency feelings over intentional binding is not possible, or (2) the present study failed to manipulate the explicit experience of agency effectively. The hypothesis that a top-down influence of higher-order explicit agency on low-level intentional binding cannot occur is plausible especially if interpreted in light of Synofzik et al.'s (2008) model, according to which the computation of an implicit feeling of agency (FoA) precedes the explicit attribution of authorship (Judgement of Agency; JoA) made when an action is categorised as not self-caused. Had the control variable in the experiment affected time estimates, it would have threatened the coherence of the model, as the two steps Synofzik and his colleagues propose would not have followed one another in a consequential fashion (i.e., explicit JoA being computed only if the implicit FoA computation generated the verdict of a not self-made action).

The possibility that the lack of an effect is due to the ineffectiveness of the manipulation presently used is also plausible, as indeed it sought to influence SoA at the most explicit level possible (i.e., by openly telling subjects “Now you can control the effect” or “Now the effect is random”). Indeed, even though participants did not seem to be monitoring their reaction time (RT), given the lack of a relationship between RT and time estimates in both high and no-control blocks, they might have consciously or unconsciously felt that the speed of their response did not affect the direction of the effect (a follow-up question assessing that after the end of the experiment would have been useful). Perhaps a more subtle manipulation of explicit control would have yielded different results, and given that, to the author's knowledge, this is the first experiment to have tried addressing this hypothesis, attempting a replication of the study using a different type of manipulation would be useful.

The marginally significant effect of control level over precision is encouraging in that respect, as it suggests that there may be some consequence of priming feelings of explicit control on some phase of the computation of low-level agency feelings. For example, it may be that telling participants the effect was random tended to slightly reduce the level of attention they paid to estimating the duration of the AEI (despite judging its length was their task), as a "not self-caused" categorisation of the action had already been made. The control variable also interacted with compatibility in affecting the precision of time estimates, with incompatible trials producing significantly less precise judgements of AEI than compatible trials only in the no-control block. If considered within this framework, the interaction could be taken to mean that action-effect compatibility served as an external cue to potential agency, reallocating attentional resources to the time estimation task, thus explaining the greater precision in compatible than incompatible trials in the no-control block but not in the high-control block, in which the necessary attentional resources had supposedly been already allocated to estimating the time interval prior to the observation of the effect because of the control priming.

Therefore, in sum, the initiation of the computation of agency may rely on the presence of certain internal cues that filter events in the environment of which an individual may be the cause. If these cues are present, then attentional resources are allocated to the computation of agency. If they are not present, then attentional resources are cut off unless an external cue (e.g., compatibility between action and effect) is present that overrides the lack of internal signals, triggering the computation of SoA. This explanation is clearly highly speculative, as it is a post-hoc conjecture based on results from an analysis conducted as a means to explore the data set at greater depth, and which yielded mostly more or less marginally significant effects. Therefore, further research is necessary to either confer a degree of credibility to this hypothesis or refute it altogether, especially since current and past studies have focused on manipulating whether one feels to be the cause of an event, rather than on what triggers the computation of that feeling in the first place. If the data were to support this explanation, it may have implications for mental health disorders in which the SoA is either enhanced or decreased, like depression and schizophrenia, shifting the focus from exclusively why they feel more or less in control of events to why they over- or under-compute SoA.

Regarding the second experimental hypothesis, the presence of a marginally significant effect of compatibility on participants' judgements of time durations suggests that even when an individual reacts rather than acts intentionally, external factors might contribute to an implicit SoA in a similar way as they have been found to contribute to it following intentional actions (Ebert & Wegner, 2010), thus causing an intentional binding effect. Therefore, it may well be that the temporal binding between action and effect does not depend on the intention to act, rather on a combination of cues, much as has been found to be the case for explicit judgements of SoA. It is hard to distinguish whether the compatibility causing the temporal binding in the present experiment was between action and effect, or perhaps between target arrow and effect, where the target arrow would have served as a supraliminal effect prime. Whereas it would be necessary to shed light on this ambiguity in future research, it does not undermine the conclusion that intentional actions do not seem to be necessary for a low-level SoA to be present. It may also be of interest to attempt replicating this effect and comparing it to what happens in situations where participants are passively induced to making a movement, thus lacking both the intention to act and the efferent motor signals generating movement, which are thought to make an important contribution to SoA (Blakemore et al., 1999), and to situations in which actions are not speeded, thus looking at binding over a scale of intentionality of action.

The results of the analysis performed on the standard deviations of participants' estimates also revealed a number of marginally significant interactions that have not been included in the discussion of the findings thus far, and which involve mostly the order of block presentation variable. Given the exploratory nature of that analysis, only results that could be readily interpreted in terms of previous findings and existing models of SoA were given consideration, as they might provide the basis for the formulation of research questions to be addressed in future studies. An interpretation of these interactions whose connection to present SoA theory was not obvious was avoided, as it would likely require both a high degree of purely post-hoc speculation and an appropriate platform for digression from the topics of the current study.

In conclusion, the present study found that explicit SoA does not affect the low-level experience of agency captured by intentional binding. A slight decrease in the precision of time estimates in trials in which participants were primed not to feel in control of an event as compared to trials in which they were primed to feel in control suggests that higher-order SoA may either influence some stage of the computation of agency or act as a trigger initiating the computational process in the first place. Further research addressing the topic is required, as this is the first study to the author's knowledge in which a downward influence of higher-order on low-level SoA is attempted. In addition, the finding of a marginal effect of compatibility between action and effect over participants' time estimates in a task using speeded reactions rather than intentional actions indicates that intentional binding may not require the *intentional* component of action to arise, if external cues suggesting agency are present. Further research is needed to support the claims presently made in the discussion of the findings, as the results obtained are mostly only marginally significant and in part the outcome of a post-hoc analysis, thus calling for particular caution in their interpretation.

**References**

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10511643

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640. doi: 10.1038/2870

Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. *Cognition*, *125*(3), 441–451. doi: 10.1016/j.cognition.2012.07.011

Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex*, *23*(5), 1031–1037. doi: 10.1093/cercor/bhs059

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. doi: 10.1016/j.concog.2009.10.002

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000a). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, *31*(2-3), 357–363. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10719163

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000b). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *355*(1404), 1771–1788. doi: 10.1098/rstb.2000.0734

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390–R392. doi: 10.1016/j.cub.2012.02.040

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385. doi: 10.1038/nn827

Hyder, F., Phelps, E. A., Wiggins, C. J., Labar, K. S., Blamire, A. M., & Shulman, R. G. (1997). “ Willed action ”: A functional MRI study of the human prefrontal cortex during a sensorimotor task. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 6989–6994.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*(3), 623–642. doi: 10.1093/brain/106.3.623

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, *18*(4), 1056–1064. doi: 10.1016/j.concog.2009.05.004

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgment of causality by human subjects. *The Quarterly Journal of Experimental Psychology*, *41B*(2), 139-159.

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., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, *96*(3), 215–231. doi: 10.1016/j.cognition.2004.08.002

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480–492.

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