Words of Agency: Executed and observed vocal actions induce a temporal binding effect

Luca Pascolini, Lisa J. Stephenson, Andrew P. Bayliss, & Natalie A. Wyer

University of East Anglia, United Kingdom

Word Count (excluding title, references, author affiliations, acknowledgments, figures and figure legends, and abstract): 9565

Abstract

Humans generally experience a sense of agency over the outcomes produced by their motor actions. This has been well established in the case of manual actions that directly affect the physical environment. Vocalisations are also actions, but they typically have only indirect effects on the environment. In the present research, we explore whether the outcomes produced by vocalisations also elicit a sense of agency. In three experiments, using an interval reproduction task, we find that performing a vocal action that produced an auditory outcome caused participants to underestimate the amount of elapsed time between actions and outcomes (i.e., temporal binding), an implicit index of the sense of agency (Experiment 1). We also show that observing others produce vocal actions elicits temporal binding, but only when the observer has direct visual access to the vocal action being executed (Experiments 2 and 3). Taken together, our findings suggest that direct observation of an action is necessary to experience a temporal binding effect for actions performed by others, and that audio-visuomotor information may play a role in the generation of temporal compression experienced over observed actions

Public Significance Statement

When we perceive that an action causes an outcome, we tend to underestimate how much time elapsed between cause and effect. This phenomenon, known as 'temporal binding' has been shown in a range of situations where physical actions (e.g., flipping a light switch) produce an outcome (e.g., a light coming on). For the first time, this research shows that a similar phenomenon occurs when something we say (rather than do) causes the outcome, and therefore provides new insights into how beliefs about causality affect the perception of time. These findings also contribute to our understanding of how agency is experienced over outcomes

produced by vocal actions in the context of interaction with human and artificial agents, promoting an efficient design of the latter.

Keywords: Sense of agency, speech, action perception, temporal binding

Words of Agency: Executed and observed vocal actions induce a temporal binding effect

- Alexa, play “The Ties That Bind” by Bruce Springsteen -

Detecting and perceiving changes to our environment is a core function of the cognitive system. These changes occur for a multitude of reasons; take for example a song starting, or a lamp lighting. Such a change could be because we acted to make it happen (i.e., I flicked the lamp switch), someone else acted to make the change (i.e., you flicked the switch), or because of factors that we cannot directly observe (i.e., the switch has a timer). Our own motor system is a frequent cause of environmental changes , but many events are caused by other agents or occurrences. Our motor system monitors action outcomes to both optimise action control, and also to correctly claim authorship over events in the environment we caused (Gallagher, 2012). This results in a sense of agency: a recognition of ourselves as agents of goal-directed actions that produce changes in the environment. An accurate sense of agency helps us to distinguish *what we are doing* from *what is happening*. Here, we explored the agency mechanisms in the communicative action domain, with specific focus on self-generated and observed social actions (e.g., vocalisations) and manual actions.

Sense of agency arises when outcomes of voluntary actions match predicted outcomes (e.g., Haggard, 2017). According to the comparator model of action control (Blakemore et al., 2002), motor programs are generated to select the appropriate action to achieve a goal. The selected motor program is then used to predict the action’s sensory outcomes. After the action, sensory information is compared with what was predicted, and may produce a prediction error if the two do not match. Typically, however, if an event is caused by one’s own action, the predicted feedback *will* match the actual sensory information, resulting in a zero prediction error. Accordingly, we experience a sense of agency over outcomes that can be predicted given their motor programs. Hence, time compression (or *temporal binding*) can be seen as an implicit marker of agency.

Temporal Binding as an Implicit Measure of Agency

The temporal binding paradigm provides a way to assess sense of agency without relying on self-reports (Haggard et al., 2002). Research employing this paradigm has demonstrated that the perceived temporal distance between an action and an outcome is compressed when the two are causally related, but is relatively accurate when the two events are believed to be independent. The temporal binding effect is theorised to measure an *implicit* dimension of the sense of agency (see Haggard, 2017, for a review) and is argued to be dissociable from explicit self-reports of agency (feeling of responsibility for an outcome), although the two are not always independent (Moore et al., 2012). Consistent with this, Synofzik and colleagues (2013) proposed an optimal cue integration account wherein implicit agency is assumed to operate at a sensorimotor level, whereas explicit agency results from higher-level reasoning.

Temporal binding has been measured using a variation of the Libet clock method (Libet et al., 1983) in which participants are asked to observe a rotating hand on a clock while executing actions and observing outcomes. Participants are asked to report where the clock-hand is positioned at the instant of the participant’s action or its outcome(Haggard et al., 2002). Participants typically report that the hand’s position is *later* for their own actions (e.g., pointing to 12 instead of 8) and *earlier* for the outcomes (e.g., pointing to 19 instead of 23). In contrast, participants report the clock-hand’s position accurately when the same events occurred independently. An adaptation of this technique, also previously used to assess temporal binding, is the interval reproduction task (Humphreys & Buehner, 2010). Here, participants are presented with a variable time interval between action and outcome, and then asked to reproduce the interval by pressing and holding down a key for the estimated duration.

Perceiving Agency in Social Contexts

Typically, the temporal binding effect has been investigated by assessing the perceived interval between a physical action (e.g., a button press) and the auditory (e.g., Engbert & Wohlschläger, 2007; Haggard & Clark, 2003; Moore et al., 2009) or visual (e.g., Cravo et al., 2011) outcome. More recently, the role of the sense of agency in social contexts (where others may be responsible for observed outcomes) has received increasing interest (e.g., Capozzi et al., 2016; Engbert et al., 2008; Stephenson et al., 2018; Strother et al., 2010; Wohlschläger et al., 2003; see Moore & Obhi, 2012, for a review). This research has explored whether one's own sense of agency is affected when acting and interacting with others, and whether a temporal binding effect can be experienced over observed actions performed by someone else.

Executing actions at the same time as someone else produces their own actions might present challenges to correctly claiming or inferring agency. When it comes to feeling control over outcomes produced in concert with others, the preponderance of evidence suggests that participants show a reliable temporal binding effect for their own actions and their outcomes – even when another actor might have plausibly been responsible for those outcomes (e.g., Engbert et al., 2007; Obhi & Hall, 2011a; Pfister et al., 2014). Such findings beg the question of whether and/or when agency is experienced in the context of observing others' actions. On one hand, people should not experience a sense of agency over actions they observe others executing. Being able to distinguish between what we control and what is controlled by others is fundamental to a coherent sense of agency as described above: a mechanism that accounts for what we are doing compared with what is happening. On the other hand, being able to distinguish whether an event has occurred spontaneously or as the result of someone else's actions, would confer adaptive benefits for individuals striving to understand cause-effect relationships within a social context. Perhaps not surprisingly, then, the evidence is mixed as to whether temporal binding occurs in relation to observed actions, with some studies detecting time compression (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003) and others not (Engbert et al., 2007; Engbert et al., 2008, Obhi & Hall, 2011b).

One possible explanation for such discrepancies is offered by considering how temporal binding for observed actions might relate to action observation more generally. Prior investigations into the mirror neuron system (Di Pellegrino et al., 1992; Gallese et al., 1996; see Rizzolatti & Craighero, 2004, for a review) suggest that observing someone else executing an action may trigger motor simulation processes, where self-generated and observed actions are computed analogously. In this way, observing an action could engage the comparator model, resulting in the generation of a temporal compression which is not associated with a concurrent explicit sense of agency.

Importantly, studies that did not detect temporal binding for observed actions employed a method where, to promote experimental control, the other's action was not visible to participants (or indeed was not actually performed by the other actor). For example, Obhi & Hall (2011b) had participants and confederates separated by a curtain, and the action was always produced by the participant. Similarly, Engbert and colleagues (2007) had participants observe an illusory action, where the hand of the experimenter was just lying on a self-depressing lever.

The mixed results of temporal binding for observed actions have been interpreted by some as suggesting that temporal compression has little to do with sense of agency, but reflects merely a causal (not necessarily intentional) relationship between two events (Buehner, 2012;

Buehner & Humphreys, 2009). Whilst we would not argue that observing others' actions generates an explicit sense of agency or feelings of control over outcomes produced by someone else, temporal binding is dissociable from explicit agency and may therefore occur in its absence. The evidence reported by previous research, showing that fake and illusory actions failed to produce a binding effect, led us to wonder whether time compression might occur only for observed actions through the activation of the mirror neuron system and would thus necessitate direct observation of the action. Given the crucial importance of action observation in the human interaction domain, we aimed to shed some light on this debate.

Mechanisms and Outcomes of Agency

Perhaps the most obvious actions deployed to directly modify the environment are manual, such as reaching, grasping, or throwing. However, our ability to influence the environment is not limited to these physical manipulations. We also execute a broad range of communicative actions (i.e., facial expressions, gaze, speech) that, while unable to directly affect the physical environment with detectable outcomes, still have the power to change the environment *indirectly* by influencing the behaviour of others. Such social actions convey a codified signal to a designated receiver, and can change other people's mental states and behaviours (e.g., Caspar et al., 2016; Stephenson et al., 2018). In this way, communicative actions produce a change from the current state of the environment to our desired one, in the same way as physical actions do. An open question, however, is the extent to which the indirect effects of non-manual actions may elicit similar signatures of the elicitation of a sense of agency. Furthermore, being able to detect the outcomes produced by other people is essential for understanding their actions and ascribing appropriate mental states (Happé et al., 2016). Hence, establishing how agency is experienced and attributed when it comes to communicative actions is crucial for our understanding of social cognition.

Of course, manual and vocal actions differ from each other, and they may not necessarily recruit the same mechanisms for generating perceptions of agency. Yet, having a strong agency detection system that accounts for both self-generated physical and social actions could be more efficient than having two separate and dedicated systems. Recent research on another type of communicative action – changing the direction of eye gaze – is suggestive. Stephenson and colleagues (2018) adopted a temporal binding paradigm to test whether initiating joint attention could lead to an implicit sense of agency towards the social outcome produced (the follower gazing at the same object). They found that leading the gaze of an on-screen face induced an underestimation of the temporal gap between action and outcome. This is consistent with Pfeiffer and colleagues (2012), who reported that participants felt more control over congruent than incongruent gaze responses induced in other people.

Alternatively, there are also reasons to speculate that sense of agency operates in different ways within physical and social domains. Indeed, we continuously experience physical actions producing immediate outcomes. The temporal gap in which we can assess whether our action was successful is very short, and leaves little to no ambiguity: if the action did not produce the intended change, it needs correction. In contrast, when considering communicative actions which may produce changes in another person, the time lag until the outcome is achieved can be much longer, making it more difficult to predict (Kunde et al., 2017). Hence, there may be much more ambiguity in assessing the effectiveness of our communicative actions, due to the larger and more variable amount of time to experience outcomes.

This difference may be a crucial one for how sense of agency functions in physical compared with social contexts. The greater variability in the delay between communicative actions and outcomes could lead to no binding effect at all, as social agency detection could be ascribed to higher-level processes such as theory of mind (Premack & Woodruff, 1978). In fact, even if no implicit marker of agency is generated (namely, subjective time compression), we could still rely on explicit higher-level mechanisms to efficiently detect whether our communicative actions affected someone's behaviour. On the other hand, given the importance of social agency detection, the greater unpredictability of communicative actions could lead to *stronger* implicit effects. Such implicit effects could rely on a flexible system, able to account for the intrinsic variability of timing in social interactions. Hence, whether vocal actions could generate a temporal binding effect (associated with implicit sense of agency) is an important question for social cognition.

Speech as an Action

Previous research provided evidence that temporal compression can be experienced when performing actions with our mouths (e.g., blowing; Yabe et al., 2019). In the present research, we sought to establish whether vocal actions (i.e., speech signals) that prompt a concrete outcome (an audible tone) give rise to temporal binding. In doing so, we attempt to isolate vocal actions from their social context. It is important to acknowledge that such contexts add complexity to the task of deciphering agency. Vocal actions in the context of social interaction produce a vast range of outcomes. They may influence others' motor actions (that is, a purely physical influence, for example, asking someone to flick a light switch). They may also affect others' social behaviour (that is, a hybrid between social and physical influence, for example, telling someone to look at something to redirect their attention). Further, they may influence others’ cognitive and emotional states (that is, a purely social influence, for example a lecturer presenting new concepts to a student, changing their knowledge states). Thus, when talking to other people, we may influence them in ways that can be directly or indirectly observed.

Furthermore, previous research has generated evidence suggesting a cognitive representational overlap between speech and action. Specifically, observing vocal actions recruited brain regions known to be involved in action perception, such as the premotor and the adjacent primary motor cortex (Andric et al, 2013; Skipper et al., 2005). Vocal actions, of course, also differ in at least one critical way – notably, they can be heard instead of, or in addition to, being seen. Thus, we investigated whether visually observing – or merely hearing – another person speak is necessary for temporal binding. This research may provide insight into the mechanisms that contribute to temporal binding for observed actions. In fact, if temporal binding effects require only recognition of a causal relationship between two events, we should expect participants to underestimate the temporal gap in all conditions where they know that an action produced an outcome (regardless of who produced it, or whether the action could be directly observed). On the other hand, if temporal binding effects are informative about intentionality through motor simulation processes, we should expect a discrepancy between vocal actions that can be genuinely observed and vocal actions that can only be heard. To our knowledge this topic has never been investigated before, and as such we aimed to directly compare physical and vocal actions producing the same kind of outcome commonly used in temporal binding paradigms (an auditory tone).

Overview of Experiments

In three experiments, we tested the hypothesis that vocal actions – like manual actions – give rise to a sense of agency, as measured implicitly by the temporal binding effect (see

Haggard, 2017, for a review). Participants' binding effects were compared across different conditions: when they performed a physical or vocal action, when they observed the execution of a physical or vocal action, and when they were witnessing two independent events. We predicted that we would find reliable temporal compression over self-generated physical actions and self-generated vocal actions. According to previous literature (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003) we also predicted that genuine and observable physical and vocal actions performed by other people would result in temporal underestimation towards their outcomes. Since we aimed to explore the agency contributions of action observation, we adopted an interval reproduction task (Humphreys & Buehner, 2010) as a paradigm to measure participants' time compression. In fact, unlike the Libet clock method, the interval reproduction task does not allow for the individual contributions of action and outcome to be assessed, but it does allow more flexibility in terms of visual stimuli presentation, since participants do not need the rotating clock to provide their answers. In all experiments detailed below, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures collected.

Experiment 1

In Experiment 1, participants completed an interval reproduction task under four different conditions, manipulated within-subjects. Each condition featured a different start stimulus, each of which produced an identical outcome. First, we sought to replicate previous findings using this task (e.g., Howard et al., 2016; Humphreys & Buehner, 2010; Poonian & Cunnington, 2013; Stephenson et al., 2018) by including an Operant Manual condition where participants reproduced the time interval between their own manual action (pressing a key on the computer keyboard) and a tone produced in response to that action. To test whether vocal actions were also capable of producing a similar level of temporal compression, we included an Operant Vocal condition where participants reproduced the time interval between a vocal action (i.e., saying the word ‘Go’) and a tone produced in response to that action. On a second point, we aimed to provide further evidence relating to the long-term debated possibility that observing actions performed by others can also produce a temporal binding effect. To do this, we included an Observed Manual condition, where participants reproduced the time interval between a manual action performed by the experimenter (i.e., the participant observed the experimenter pressing a key on the computer keyboard) and a tone produced in response to his action. This condition provides a conceptual replication of previous studies using an interval reproduction task in lieu of the Libet clock method (e.g. Wohlschläger et al., 2003). As is typical for temporal binding paradigms, we compared performance in the three experimental conditions (Operant Manual, Operant Vocal, and Observed Manual) with a Control condition in which no action was performed by any agent. Here, participants reproduced the time interval between two tones that were automatically produced by a computer. As a manipulation check, we also asked participants to report how much they felt in control of the action outcome in the different conditions, providing an explicit index of their sense of agency.

Methods

Participants

A power analysis (carried out with G*Power 3; Faul et al., 2007) indicated that to detect a medium-large effect size (as reported by Stephenson et al., 2018 for a direct comparison between two experimental conditions) $d_z = 0.69$, with $1 - \beta = 0.95$ at $\alpha = 0.05$, a minimum sample size of 30 would be required. Therefore, 32 participants (a sample size capable of detecting an effect size $d_z = 0.66$, with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 18-33 years ($M = 20.24$, $SD = 2.65$, 28 were females), recruited from the University of East Anglia, completed the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported having normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Apparatus

The experimental setting consisted of two adjacent chairs and a table in a dimly lit room. A computer monitor (BENQ XL2411: size: 24”; resolution: 1920x1080; refresh rate: 60Hz) and a set of external speakers (Bose Companion 2 Series III) were placed on the table and used to display experimental stimuli. A Chronos multifunctional response and stimulus device (Psychology Software Tools, Inc., Sharpsburg, PA, USA) with microphone, and an external keyboard (Kensington KP400) were placed on the table and used to collect participants' responses. The height of the table was 80cm. The position of the monitor was centred to participant's body midline, 60cm from the edge of the table. The position of the speakers was respectively 30cm on the right (right speaker) and 30cm on the left (left speaker) of participants' body midline, 50cm from the edge of the table. The position of the keyboard was centred to participants' body midline, 25cm from the edge of the table. The position of the response box was 40cm on the right of participants' body midline, 25cm from the front edge of the table. The position of the microphone (only used in the Operant Vocal condition) was centred to participants' body midline, 5cm from the edge of the table.

Stimuli

A first low-pitch tone (150ms, 440Hz sine wave, sample rate: 44100Hz, bit-rate 16: Poonian & Cunington, 2013) was created as the start stimulus in the Control condition. A second high-pitch tone (100ms, 1 KHz sine wave, sample rate: 44100Hz, bit-rate 16: Humphreys & Buehner, 2010) was created as the end stimulus in the Control condition and as the action's outcome in the Operant Manual, Operant Vocal and Observed Manual conditions. The experiment was run using E-prime version 3.0 (Psychology Software Tools, Inc., Sharpsburg, PA, USA). All auditory stimuli were created using MATLAB (MathWorks, Natick, MA, USA).

Design

The temporal binding effect was measured using an interval reproduction task (Humphreys & Buehner, 2010), where participants were asked to reproduce the duration of the time interval between two events by pressing and holding down the central key on a response box for the same amount of time. To effectively manipulate participants' agency perceptions, the nature of the first event varied in each block, while the second event remained unaltered.

Procedure

Participants were invited to the laboratory and welcomed. They read an information sheet and provided informed consent to join the study. Prior to beginning the experiment, participants were introduced to the experimental equipment. They completed a sample of ten experimental trials (five Operant Manual trials and five Operant Vocal trials, without the reproduction task) set to demonstrate that when they executed an action (whether manual or vocal), a high tone would be released. This served to ensure that they understood the causal relationship between the two events.

Participants completed four blocks of trials each featuring a different start stimulus. Each block consisted of 5 practice trials and 30 experimental trials, for a total of 120 experimental trials. Block order was pseudorandomized across participants using the latin square method. Control trials began with a white fixation cross displayed in the centre of the screen (1500ms) after which, at a random interval (1500-2000ms) a low-pitch tone (150ms, 440Hz) was presented. Then, after the target interval of time (randomised between 500 and 1500ms, on a continuous level) a high-pitch tone (100ms, 1kHz) was presented. Participants were then asked to reproduce the duration of the interval between the two tones by pressing and holding down the central key on the response box for the same amount of time. After participants completed the reproduction task for the considered trial, they were asked to press the space bar to begin the following trial.

In the Operant Manual trials, participants were instructed that after the fixation cross disappeared (1500ms) performing a specific action (i.e., press the 0 key on the number pad at any moment of their choosing) would produce, after a variable time interval (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz). Participants were encouraged to avoid rushing in giving the action, or to start the action at any predetermined fixed time (e.g., counting to 3 and perform the action). Following the tone participants were instructed to press and hold down the central key on the response box for a duration equal to that of the time interval between their action and the tone. The target time interval was set to begin after the end of participants' action (i.e., when the 0 key on the number pad was fully released). After participants completed the reproduction task for the considered trial, they were asked to press the space bar to begin the following trial.

The Operant Vocal trial procedure was identical to Operant Manual trial procedure, except for the action that participants were asked to perform. Participants were instructed that in the Operant Vocal trials, performing a specific action (i.e., pronouncing the word "go" in the microphone) would produce, after a variable time interval (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz). Participants then completed the reproduction task, pressing and holding down the central key on the response box for a duration equal to that of the time interval between their action and the tone. The target time interval was set to begin after the end of participants' action (i.e., when the microphone's noise threshold reached 1%). After participants completed the reproduction task for the considered trial, they were asked to press the space bar to begin the following trial.

Using the same trial procedure of the Operant Manual condition, participants completed the Observed Manual condition where they observed the experimenter performing an action (i.e., pressing the 0 key on the number pad at any moment of his choosing). After a variable period of time (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was produced. Participants then completed the reproduction task and were asked to press the space bar to begin the following trial.

After each block of trials, participants self-reported the degree to which they felt control over the high-pitch tone. The instruction on screen was "Please rate 1 to 8 how much control you felt over the high tone, 1 meaning no control and 8 meaning a lot of control" (Beyer et al., 2018). Graphic representation of the trial procedure is shown in Figure 1.

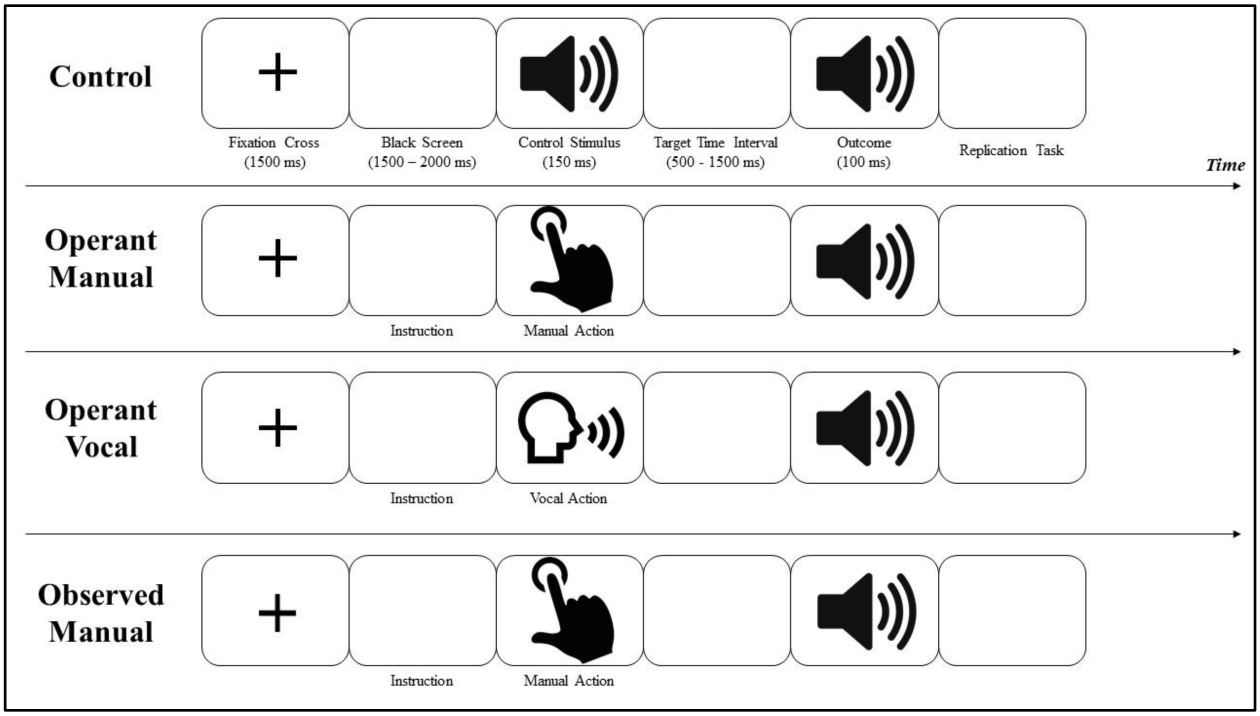


Figure 1. Trial Procedure for the four conditions in Experiment 1.

When participants were completing the Control, Operant Manual, and Operant Vocal conditions, the experimenter sat approximately 120cm distant from the participants' position, on the orthogonal side of the table. During the Observed Manual condition, the experimenter sat adjacently on the right of the participants, approximately 10cm distant from their position.

Raw data and experimental stimuli used in this study has been made public to comply with good open science procedure. All files can be accessed at https://osf.io/jp57r/. This study was not preregistered.

Results

The calculated dependant measure was the inter-event proportional interval reproduction (Stephenson et al., 2018), derived by dividing the reproduced time interval by the actual time interval for the same trial (ms). Thus, scores equal to 1 represented perfect accuracy, while scores greater than 1 were over-reproductions, and scores lower than 1 were under-reproductions (that is, subjective temporal compression).

The following exclusion criteria were decided prior to the data collection: participants were excluded if they failed to produce temporal intervals covarying monotonically with actual action-tone intervals (average ρ across conditions lower than 0.4, meaning that participants were not continuously putting an effort in following the instructions; Caspar et al., 2016). Individual trial data were excluded based on failure to respond (trials where the reproduced time interval was lower than 100ms were discarded, meaning they had just pressed the key rather than pressed and held it) or extreme variability: trials falling outside +/- 3 SD from individual mean were removed, as is usual practice in temporal binding research using interval reproduction (e.g., Howard et al., 2016; Stephenson et al., 2018). This resulted in 3 participants excluded and 22 trials removed (0.57% of total trials).

Mean proportional interval reproductions were calculated for each participant on each condition, and submitted for statistical analysis. A one-way within-subjects ANOVA was conducted on proportional interval reproduction scores. There was a significant effect of the type of action: $F(3, 84) = 7.438$, $p < .001$, $\eta_p^2 = 0.210$. According to our initial experimental hypotheses, planned paired-samples t-tests revealed that the mean proportional interval reproduction in the Control condition (95% *CI* [0.94, 1.15]) was significantly higher either than the Operant Vocal condition (95% *CI* [0.88, 1.05], $p = .026$, $d_z = 0.44$), the Operant Manual condition (95% *CI* [0.84, 0.99], $p < .001$, $d_z = 0.80$), and the Observed Manual condition (95% *CI* [0.85, 1.00], $p = .003$, $d_z = 0.63$). No significant differences were found between the Operant Manual, Operant Vocal, and Observed Manual conditions: $F(2, 56) = 1.815$, $p = .172$, $\eta_p^2 = 0.061$. Figure 2A shows a graphic representation of the data. Mean proportional interval replications, SD, and relative pairwise comparisons across all experiments are summarised in Table 1.

When comparing self-reported explicit agency ratings, planned paired-samples t-tests revealed that the mean explicit agency rating in the Control condition did not differ from the Observed Manual condition, but was lower than both the Operant Manual condition ($t(28) = -5.741$, $p < .001$, $d_z = 1.08$) and the Operant Vocal condition ($t(28) = -4.980$, $p < .001$, $d_z = 0.94$), which did not differ from each other.

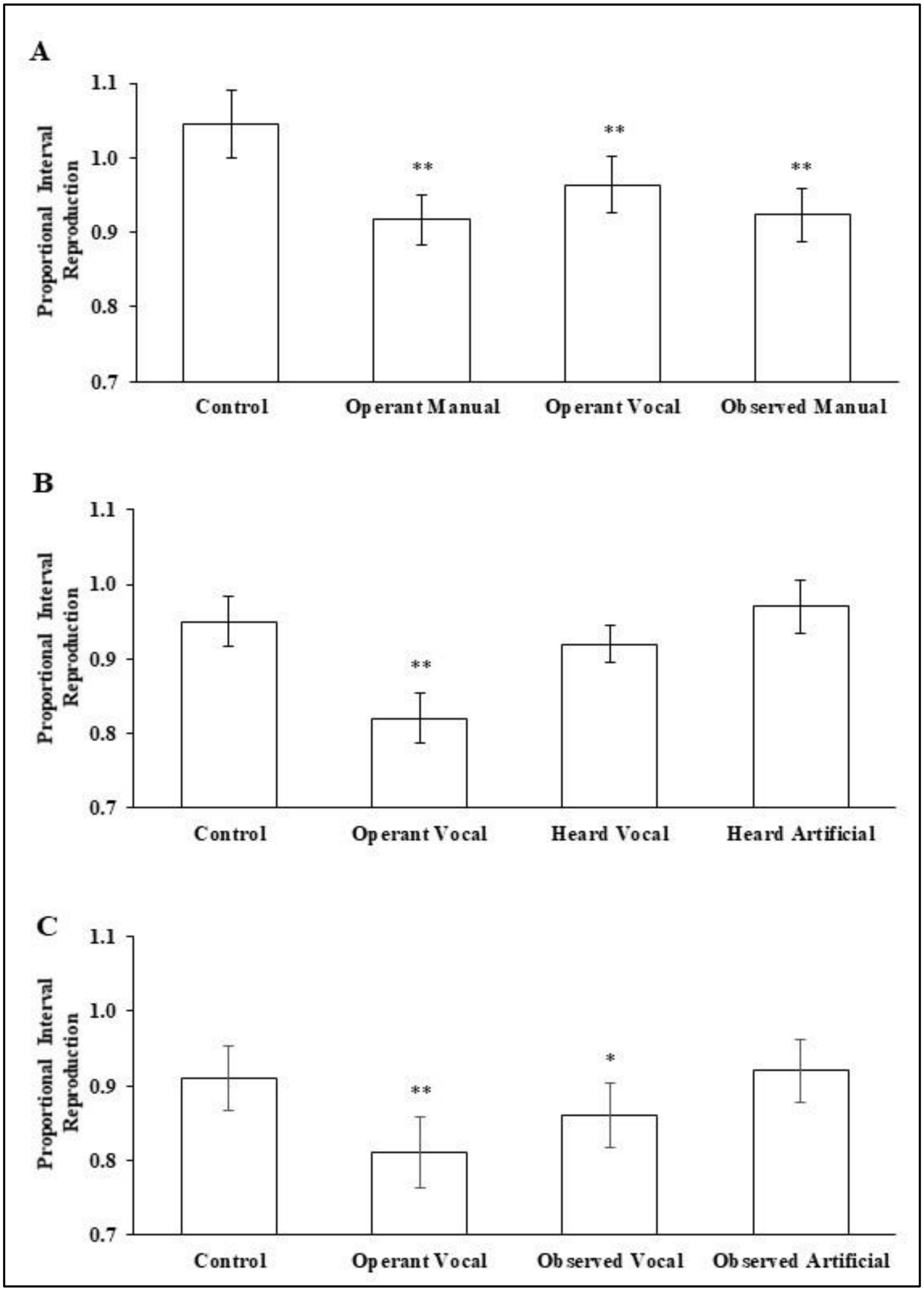

A
Proportional Interval Reproduction
1.1
1.0
0.9
0.8
0.7
**
**
**
Control
Operant Manual
Operant Vocal
Observed Manual
B
Proportional Interval Reproduction
1.1
1.0
0.9
0.8
0.7
**
Control
Operant Vocal
Heard Vocal
Heard Artificial
C
Proportional Interval Reproduction
1.1
1.0
0.9
0.8
0.7
**
*
Control
Operant Vocal
Observed Vocal
Observed Artificial

Figure 2. Mean proportional interval reproduction across the four conditions in Experiment 1 (A), 2 (B), and 3 (C). Error bars represent the standard error of the mean. Significant differences from the Control condition are highlighted for the 0.05 (*) and 0.01 (**) level.

Table 1

Exp 1	**Control**	**Operant Manual**	**Operant Vocal**	**Observed Manual**
	1.05 (0.26)[a]	0.92 (0.19)[b]	0.96 (0.21)[b]	0.92 (0.20)[b]
Exp 2	**Control**	**Operant Vocal**	**Heard Vocal**	**Heard Artificial**
	0.95 (0.19)[a]	0.82 (0.19)[b]	0.92 (0.14)[a]	0.97 (0.20)[a]
Exp 3	**Control**	**Operant Vocal**	**Observed Vocal**	**Observed Artificial**
	0.91 (0.24)[a]	0.81 (0.27)[b]	0.86 (0.24)[b]	0.92 (0.24)[a]

Table 1. Mean Proportional Interval Reproductions for each condition across all experiments. Respective standard deviations are reported in brackets. Different superscripts within the same row indicate a significant difference between conditions ($p < 0.05$).

Discussion

The data show that participants consistently under-reproduced the temporal interval between the action and its subsequent outcome, both when they were executing a manual action and a vocal action. This provides new evidence that individuals' vocal actions may produce an implicit sense of agency, the magnitude of which compares to other manual actions that are capable of causing changes in the physical environment. We also found evidence in favour of the debated finding that observed (manual) actions also produce subjective temporal compression. In fact, our data highlights a binding effect over observed actions and consequent outcomes, to a similar extent as for self-produced actions.

Still, the mechanism underlying this effect for observed actions is not entirely clear, and further investigations should specifically address the motor simulation processes potentially involved. Explicit agency ratings reflected the pattern suggested by implicit measures, but only for self-generated actions. On the contrary, when considering observed actions participants reported low ratings of control towards the outcome, whilst still showing time compression between the events. This finding shows that participants felt a degree of explicit agency in the Operant Manual and Operant Vocal conditions, supporting our inference that the temporal binding effect presented here reflects a dimension of sense of agency, whilst also providing further evidence that while explicit and implicit dimensions of agency may be linked, they likely do not rely on the same mechanism (Moore et al., 2012; Synofzik et al., 2013).

Experiment 2

In Experiment 2, we aimed to replicate the novel temporal binding effect for self-generated vocal actions. Furthermore, building on our finding in Experiment 1 that time compression occurred between observed physical actions and their outcomes, we tested whether *heard* vocal actions would lead to the same effect. In fact, if all actions are computed analogously (e.g., by ascertaining cause-effect relationships), we should also expect to find subjective time compression for heard vocal actions produced by another agent. To do this, we amended the experimental design used in Experiment 1 while maintaining its structure. Thus, in Experiment 2 we included the same Control and Operant Vocal conditions from Experiment 1. Furthermore, we aimed to translate the observed action condition from the physical domain (Experiment 1) to the social domain (Experiment 2). To do this, we included a Heard Vocal condition where the same communicative action was performed by the experimenter. To explore whether the capacity for intentional action was a necessary ingredient for generating binding over others' vocal actions, we also included a Heard Artificial condition where the same utterance was provided by a computer.

With this study we aimed to contribute to the long debated question of what elements generate binding effects. For example, according to Buehner and Humphreys (2009), a causal relationship between two events is sufficient to produce temporal compression (account referred as “causal binding”). On the other hand, Haggard and Chambon (2012) propose that causality is necessary but not sufficient to achieve temporal compression, while intentional action planning is needed (account referred to as “intentional binding”). If causality is the only route to temporal compression, we should expect to observe binding effects in all conditions, except for the Control condition. Alternatively, if temporal compression for observed actions requires representing those actions in our own motor system, we should not necessarily expect to find binding effects in the Heard Vocal and Heard Artificial conditions.

Method

Participants

To maintain consistency between the experiments, we recruited a new sample of equal size to Experiment 1. Therefore, 32 participants (a sample size capable of detecting an effect size $d_z = 0.66$, with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 18-30 years ($M = 19.90$, $SD = 2.12$, 25 were females) recruited from the University of East Anglia, completed the experiment. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported to have normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Apparatus and Stimuli

The experimental setting and apparatus remained unaltered from Experiment 1. Two auditory start signals were created to be used in Experiment 2. First, to create comparable auditory stimuli across all conditions, 30 audio tracks of the experimenter saying the word ‘go’ were recorded. Durations and frequencies of these utterances were analysed. The average duration was 346.72ms (*SD* = 9.26), while the average frequency was 166.87Hz (*SD* = 3.41). Hence, for the Heard Artificial condition, a voice track was created using an on-line vocal synthesizer (www.cepstral.com), and then edited to match average duration and frequency of the human voice previously analysed (347ms, 167Hz). For the Control condition, a matching tone (347ms, 167Hz) was created as start stimulus. All auditory stimuli were created, analysed or edited using MATLAB.

Design and Procedure

The within-subjects design was identical to that of Experiment 1, with the exception that the start signal was replaced in two of the four blocked conditions. The four experimental conditions were Control, Operant Vocal, Heard Vocal, and Heard Artificial. Block order was pseudorandomized across participants using the latin square method. Experimental procedure remained unaltered from Experiment 1 for the Control and Operant Vocal condition. The Heard Vocal and Heard Artificial conditions were identical to the Operant Vocal condition, except for the source of the vocal input. In the Heard Vocal condition the experimenter used the microphone to provide a vocal action (i.e., pronouncing the word “go” in the microphone), whilst remaining sat beside the participant (therefore, action execution information was not accessible). In the Heard Artificial condition the microphone was positioned towards the left speaker, and the signal provided (347ms, 167Hz) served as vocal input. In this manner, the causal relationship between the events in the Heard Artificial condition was preserved. Across all conditions, the outcome tone (100ms, 1 kHz) was played by the right speaker only. Graphic representation of the trial procedure is shown in Figure 3.

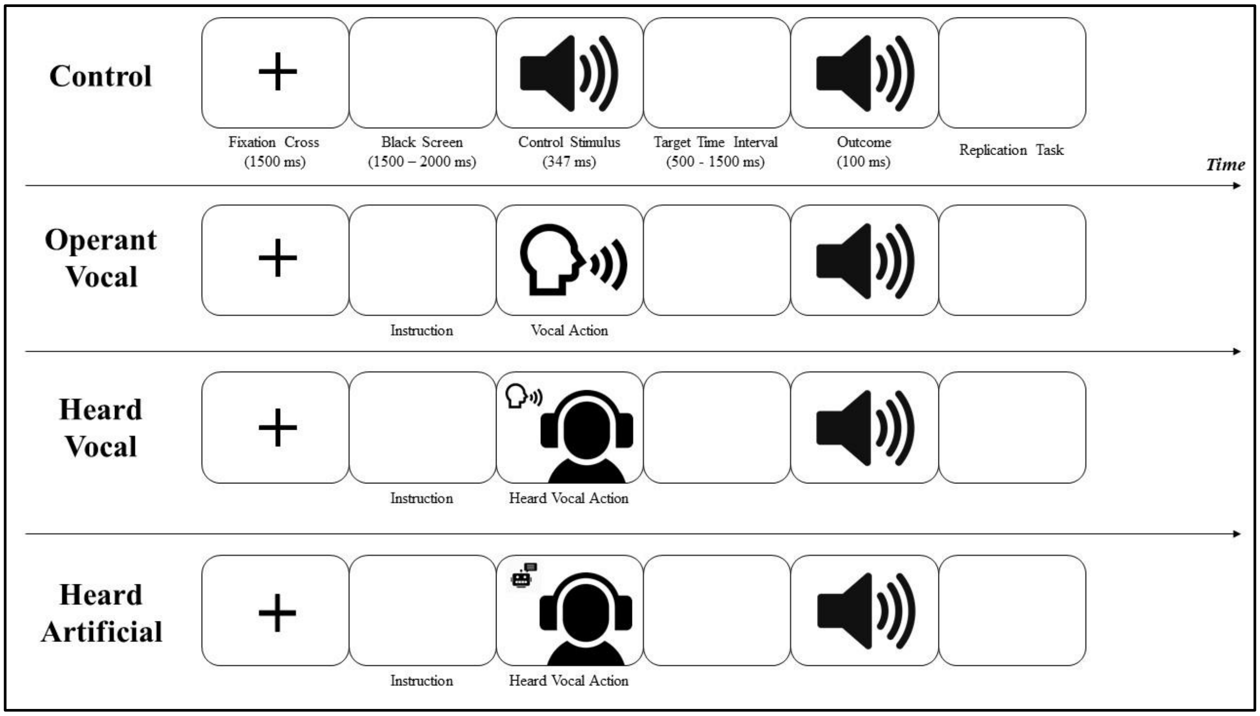


Figure 3. Trial Procedure for the four conditions in Experiment 2.

Results

The same exclusion criteria of Experiment 1 were adopted for Experiment 2. This resulted in 2 participants excluded for inaccurate reproductions and 29 trials removed (0.76% of total trials).

Mean proportional interval reproductions were calculated for each participant on each condition, and submitted for statistical analysis. A one-way within-subjects ANOVA was conducted on proportional interval reproduction scores. There was a significant effect of the type of action: $F(3, 87) = 7.757$, $p < .001$, $\eta_p^2 = 0.211$. Planned paired-samples t-tests revealed that the mean proportional interval reproduction in the Operant Vocal condition (95% *CI* [0.75, 0.89]) was significantly lower than the Control condition (95% *CI* [0.88, 1.02], $p = .001$, $d_z = 0.71$). No significant differences were found between the Control, Heard Vocal (95% *CI* [0.87, 0.97]) and

Heard Artificial (95% *CI* [0.89, 1.04]) conditions: $F(2, 58) = 1.158$, $p = .321$, $\eta_p^2 = 0.085$. Figure 2B shows a graphic representation of the data. Mean proportional interval replications, SD, and relative pairwise comparisons across all experiments are summarised in Table 1.

When comparing self-reported explicit agency ratings, planned paired-samples t-tests revealed that the mean explicit agency rating in the Operant Vocal condition was higher than the Control condition ($t(29) = 4.754$, $p < .001$, $d_z = 0.88$), the Heard Vocal condition ($t(29) = 4.882$, $p < .001$, $d_z = 0.91$), and the Heard Artificial condition ($t(29) = 5.559$, $p < .001$, $d_z = 1.03$).

Discussion

Participants again consistently under-reproduced the temporal interval between their self-generated vocal actions and their subsequent outcomes. This corroborates findings from Experiment 1, providing us with confidence over the reliability of this novel effect. Given that we also found binding for observed physical actions in Experiment 1, we expected that participants might analogously underestimate the temporal gap between *heard vocal* actions and their consequent outcomes. Importantly, no reliable time compression was observed when participants listened to a vocal action performed by the experimenter, nor when the same utterance was generated by the computer. This seems to contrast with Experiment 1, where participants showed a temporal binding effect over observed physical actions.

A possible explanation for this discrepancy is that merely *hearing* the experimenter's vocal action did not allow participants to simulate it in their own motor systems. This element drove us to investigate the effects of direct action observation on the generation of temporal compression. Indeed, the Observed Manual (Experiment 1) and Heard Vocal (Experiment 2) conditions were not comparable in the extent to which they allowed visual access to action kinematic information. Kinematics in action observation has been shown to be of critical importance to understand the action itself, and it is also associated with higher-level mechanisms such as intention attribution and theory of mind (Aglioti et al., 2008; Cavallo et al., 2016). Thus, we wondered whether direct action observation (intended as observed muscular activation) could be a contributing feature to the generation of a temporal binding effect over observed actions performed by others, which is investigated in Experiment 3.

It is worth mentioning that the data reported in Experiment 2 provides insight into the fundamental elements that generate temporal compression. Some accounts claim that temporal compression is achieved whenever a causal relationship between events is detected, whereas others imply that intentional action planning is needed. Here, in both the Heard Vocal and Heard Artificial conditions, the events were not mutually independent (as opposed to the Control condition), but linked in a cause-effect relationship. As we did not find reliable time compression in these two conditions, this argues against the view that mere causation between events is sufficient to generate a temporal binding effect. We can therefore infer that, while causation between the events is indeed a necessary prerequisite to achieve a binding effect, it may not be sufficient by itself.

Experiment 3

Experiment 2 did not show reliable time compression between heard vocal actions performed by the experimenter and their outcomes, in contrast with Experiment 1 where binding did occur for observed physical actions. Considering that self-generated vocal actions appear to be consistently capable of producing a temporal binding effect (as reported in Experiment 1 and 2), we set out to investigate the discrepancy between physical and vocal actions performed by others. The main difference between observed conditions in Experiment 1 and 2 was that while physical actions performed by the experimenter were entirely visible to the participants, this was not true for heard vocal actions. In other words, while there were two major differences between the Observed Manual and the Heard Vocal conditions, we only manipulated one of them. Namely, while participants could see the hand moving in the Observed Manual condition (Experiment 1), they could not see the mouth moving in the Heard Vocal condition (Experiment 2). Crucially, having visual access to action kinematic information is widely considered to play a major role in action prediction (Cavallo et al., 2016). Thus, if visual information plays a role in the generation of the temporal binding effect for observed actions, it could be argued that a visuomotor component (and as such an action prediction process) should be considered for extending the comparator model, as efferent copies may be generated not only from one's own motor program, but also derived from observed actions. To address this divergence, in Experiment 3 we added a visual component in every condition.

Method

Participants

To maintain consistency between the experiments, we recruited a new sample of equal size to Experiment 1 and 2. Therefore, 32 participants (a sample size capable of detecting an effect size $d_z = 0.66$, with $1 - \beta = 0.95$ at $\alpha = 0.05$) aged 18-24 years ($M = 19.35$, $SD = 1.41$, 25 were females) recruited from the University of East Anglia, completed the experiment. Sample size was determined in the same way as described in Experiment 1 and 2. Participants gave written informed consent prior to the experiment, were naïve regarding the research questions, and received course credits for their involvement. All participants reported to have normal or corrected-to-normal vision and hearing. The study was approved by the School of Psychology Research Ethics Committee, University of East Anglia.

Apparatus and Stimuli

The experimental setting and apparatus remained unaltered from Experiment 1 and Experiment 2, except for the experimenter's position in the Observed Human condition, which was now facing the participants' position. Experimenter's position was located 60cm on the left of participants, 60cm from the edge of the table (thus creating a 45° angle facing North-West from participants' midline). Two audio-visual start signals were created to be used in experiment 3 (.avi format, 25 frames/s), both composed of two video clips. For the Control condition, the audio-visual start signal consisted of a first video clip depicting a white dot on a black screen (duration randomized between 1500 and 2000ms), and a second video clip showing the white dot enlarging at a constant speed while the same tone used in the Control condition in Experiment 2 was played (347ms, 167Hz). For the Observed Artificial condition, the audio-visual start signal consisted of a first video clip displaying a static frontal medium close-up of a human avatar, created using an online avatar generator (www.voki.com) with duration randomized between 1500-2000ms. A second video clip was created, displaying the avatar opening its mouth while the same artificial voice used in the Heard Artificial condition in Experiment 2 was played (347ms, 167Hz). All audio-visual stimuli were created and edited using Adobe Premiere Pro CS6 (Adobe Systems Software Ltd, Dublin, Ireland).

Design and Procedure

The design was identical to Experiment 2, with the exception that three of the blocks presented modified start signals. The four experimental conditions were: Control, Operant Vocal, Observed Vocal and Observed Artificial. Block order was pseudorandomized across participants using the latin square method. The experimental procedure for the Operant Vocal condition remained unaltered from Experiments 1 and 2.

Control trials began with a white fixation cross displayed in the centre of the screen (1500 ms) after which both video clips composing the control audio-visual start signal were presented in sequence. During the stimulus presentation, only the left speaker was operative. Then, after the target time interval (randomised between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was presented through the right speaker. Participants were then asked to complete the reproduction task.

The Observed Vocal trial procedure was identical to the Heard Vocal trial procedure in Experiment 2, but participants could now directly observe the experimenter as he was executing the vocal action, as he was sitting in front of them. Participants were instructed to wait for the fixation cross to disappear (1500ms) and then to move their gaze towards the centre of the experimenter's face, who would then execute the vocal action at a time of his choosing. Participants were instructed to move their gaze back to the monitor immediately after the end of the observed vocal action. After a variable time interval (randomized between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was released through the right speaker, and participants completed the reproduction task.

The Observed Artificial trial procedure was designed on the basis of the Heard Artificial condition in Experiment 2. Trials in this condition began with a white fixation cross (1500ms) after which both video clips composing the avatar audio-visual start signal were presented in sequence. The microphone was positioned toward the left speaker, and the signal provided (347ms, 167Hz) served as vocal input. In this manner, causal relationship between the events in the Observed Artificial condition was preserved. Then, after the target time interval (randomised between 500 and 1500ms) a high-pitch tone (100ms, 1kHz) was presented through the right speaker. Participants were then asked to complete the reproduction task. Graphic representation of the trial procedure is shown in Figure 4.

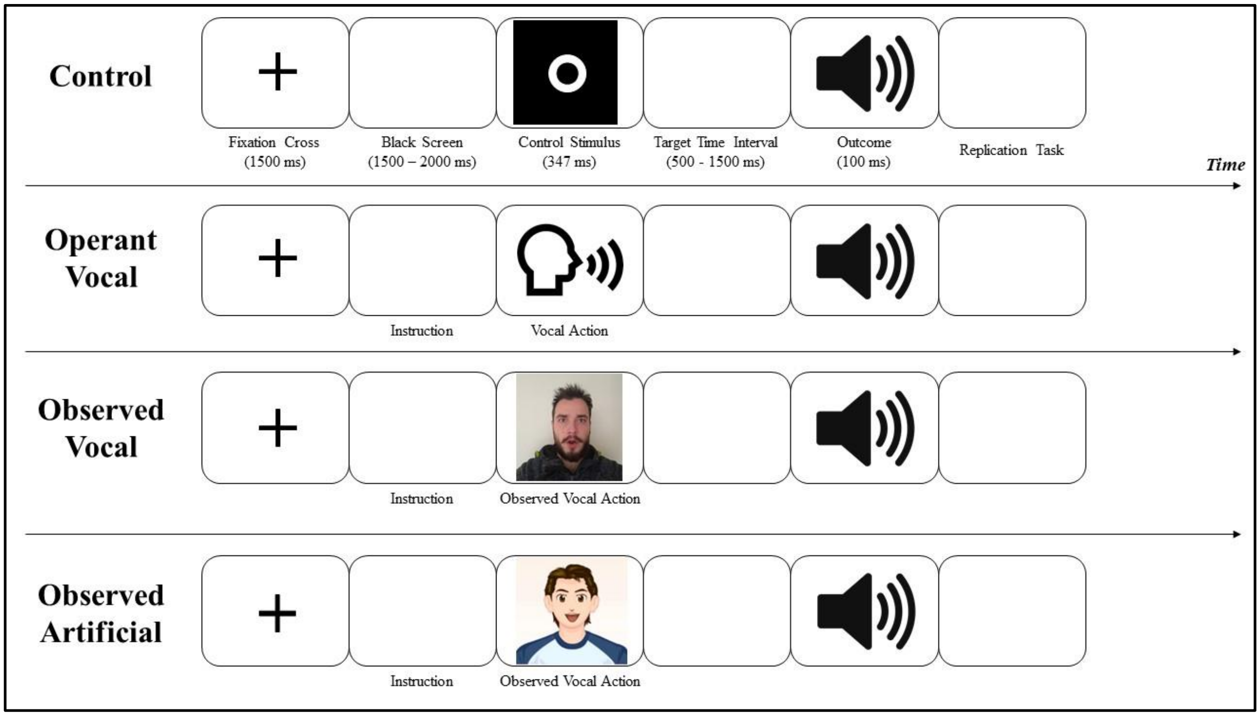


Figure 4. Trial Procedure for the four conditions in Experiment 3.

Results

The same exclusion criteria of Experiments 1 and 2 were adopted for Experiment 3. This resulted in one participant being excluded for inaccurate reproductions and 33 trials removed (0.86% of total trials).

Mean proportional interval reproductions were calculated for each participant on each condition, and submitted for statistical analysis. A one-way within-subjects ANOVA was conducted on proportional interval reproduction scores. There was a significant effect of the type of action: $F(3, 90) = 6.783$, $p < .001$, $\eta_p^2 = 0.184$. Planned paired-samples t-tests revealed that there was no significant difference between the mean proportional interval reproduction in the

Control condition (95% *CI* [0.82, 1.00]) and the Observed Artificial condition (95% *CI* [0.83, 1.01]). Contrastingly, the proportional interval reproduction in the Control condition was significantly higher either than the Observed Vocal condition (95% *CI* [0.76, 0.95], $p = .036$, $d_z = 0.40$) and the Operant Vocal condition (95% *CI* [0.71, 0.91] $p = .004$, $d_z = 0.56$), which did not differ from each other. Figure 2C shows a graphic representation of the data. Mean proportional interval replications, SD, and relative pairwise comparisons across all experiments are summarised in Table 1.

When comparing self-reported explicit agency ratings, planned paired-samples t-tests revealed that the mean explicit agency rating in the Operant Vocal condition was higher either than the Control condition ($t(30) = 5.517$, $p < .001$, $d_z = 0.98$), the Observed Vocal condition ($t(30) = 4.790$, $p < .001$, $d_z = 0.92$), and the Observed Artificial condition ($t(30) = 5.700$, $p < .001$, $d_z = 1.01$).

Discussion

Replicating our first two experiments, participants again produced a temporal binding effect for self-generated vocal actions, providing further evidence that – like manual actions – vocal actions can generate a sense of agency towards their outcome. Also consistent with our previous studies, binding was not observed when participants were shown a neutral stimulus on screen (Control condition), nor when an artificial avatar was presented (Observed Artificial condition). Strikingly however, and in contrast to Experiment 2 where participants did not show a binding effect for heard vocal actions performed by the experimenter ($d_z = 0.16$), in Experiment 3 participants showed a reliable time compression over vocal actions performed by a visible experimenter ($d_z = 0.40$). This finding seems to indicate that direct observation of an actual action is necessary to experience a temporal binding effect for actions performed by others. Taken together, our findings are consistent with the possibility that the temporal binding effects for observed actions rely on participants’ ability to represent, or embody, the actions of others.

The role of embodiment in producing temporal binding for observed actions has not received extensive investigation, and previous studies have directly compared binding for others’ actions under conditions where the action can or cannot be directly observed. Here, too, we can only draw inferences from comparing binding effects across experiments, which varied largely (0.16 vs. 0.40) despite the same original sample size. Thus, further research is needed to replicate and verify the importance of direct action observation. However, it is note-worthy that no binding emerged in the Observed Artificial condition (nor in the Heard Artificial condition of Experiment 2), providing further evidence that merely recognizing a causal relationship between events, while necessary, may not be sufficient to produce a binding effect.

General Discussion

The aim of this work was to investigate whether vocal actions were capable of producing a subjective time compression towards their outcomes. This effect, known as temporal binding, has been interpreted by some (e.g., Haggard et al., 2002) as an index of sense of agency, and by others (e.g., Buehner, 2012; Buehner & Humphreys, 2009) as reflecting the recognition of cause-effect relations between events, even in the absence of intentionality. We aimed to contribute to this debate by directly comparing conditions where causality occurred with or without intentionality. Throughout the three experiments we reported here, we obtained consistent evidence for the novel finding that vocal utterances produced a temporal binding effect, in a similar fashion as do other motor actions (e.g., hand movements). This is an indication that we feel a sense of agency for vocal actions as well, despite their inability to directly affect the physical environment. One critical implication of this is that the social environment – like its physical counterpart – affords opportunities to modify our surroundings and to experience a sense of agency over the effects we produce in other people. Our data provide evidence that a common mechanism may be accounting for the sense of agency both for physical and communicative actions, as originally theorised by Stephenson and colleagues (2018). The fact that vocal actions can produce a binding effect meaningfully expands our understanding of how agency is experienced, as most previous studies focused on outcomes produced by button presses (see Moore & Obhi, 2012, for a review).

Mechanisms of Temporal Binding

In Experiment 1 we found evidence for a temporal binding effect following the observation of others' physical actions, which is consistent with previous research (Poonian & Cunnington, 2013; Strother et al., 2010; Wohlschläger et al., 2003). Still, there is debate over why observed actions lead to temporal compression. Some authors favour the interpretation that this reflects a vicarious sense of agency, suggesting that such a mechanism could be an important feature of the empathic experience towards others' actions (Wegner et al., 2004). Their claim is that the emotional sensitivity gained for actions performed by others might be due to our ability to build foreknowledge of their actions, leading us to experience those actions as they belong to us and are under our personal control. Our data contrasts with this hypothesis: while our participants did produce a binding effect over observed actions, they did not perceive personal control over the outcome produced by those actions. This also suggests that implicit and explicit sense of agency may be driven by separate mechanisms (Moore et al., 2012).

An alternative explanation for temporal binding in action observation is that perceivers mentally represent others' actions within their own motor systems. Viewing actions performed by other individuals activates frontal and parietal cortical areas typically involved in action planning and execution (Di Pellegrino et al., 1992; Fogassi et al., 2005; Gallese et al., 1996). This process may include not only a simulation of motor planning and execution, but also a prediction of the outcomes that would be generated by the observed action (Aglioti et al., 2008). In other words, when looking at someone else's action, our Mirror Neuron System may simulate that action as if it was our own, and generate a sensory prediction of the anticipated outcome. These predictions are then processed by the comparator and matched with actual sensory feedback. If the prediction matches the actual outcome, subjective time compression would be produced following the same mechanism as for self-generated actions (Blakemore et al., 2002; Haggard, 2017). This alternative explanation is corroborated by the discrepancy we found between the heard and observed vocal actions in Experiments 2 and 3. In fact, in Experiment 2 vocal actions produced by an experimenter seated out of participants' sight did not generate a binding effect over their outcomes. However, when participants were granted visual access to the action kinematics of the experimenter's speech (Experiment 3), subjective time compression re-emerged. It is worth noting that, differently from Experiment 2, in Experiment 3 participants were instructed to shift their gaze between the experimenter's face and the PC monitor. Saccadic eye movement has been showed to produce temporal compression and induce temporal binding (Morrone et al., 2005; Yabe & Goodale, 2015). However, we note that while previous research into saccadic eye movements is important and relevant to consider in relation to our work, there are sufficient differences between prior paradigms and our own that it seems unlikely that saccadic eye movements are responsible for the binding effect we reported. In particular, posing a difference from previous research where perceived time intervals were measured between two visual stimuli, our own research focused on the time interval between two auditory stimuli, where eye movements are considerably less (if at all) relevant.

Across the three experiments reported here, auditory signals alone (whether computer-, human- or avatar-generated) were not capable of producing a temporal binding effect. This suggests that explicit knowledge of causality (by itself) may be necessary (Buehner, 2012) but not sufficient to induce an implicit sense of agency. This evidence is in contrast with previous research which advanced the idea of a causal binding to occur between two events whenever the latter is thought to depend on the prior (Buehner & Humphreys, 2009). However, methodological differences must be noted: while our task asked participants to reproduce a time interval they just experienced, Buehner and Humphreys (2009) asked participants to indicate the time point at which they predicted an outcome that they had (or had not) caused to occur. In other words, our paradigm involved a retrospective assessment of time, while that of Buehner and Humphreys used a prospective assessment task. We suggest future research could tackle this controversy, by assessing causality and intentionality as features of the temporal binding effect using a consistent methodology to allow direct comparisons.

As we predicted, our data showed no temporal binding effect for artificial vocal actions, both when they were heard and observed. This finding corroborates the idea that causality alone is not the only key to achieve temporal compression. In fact, the mere communicative act (pronouncing a word), was, when uttered by artificial agents, ineffective at generating binding effects. In other words, unlike observing another person producing a vocalization, viewing an artificial agent pronounce the same vocalization did not yield a temporal binding effect. One potential constraint on our interpretation of these results is that the artificial agent was a 2D avatar created through the animation of a drawn human face, and as such it was intrinsically different from the (3D, realistic) embodied agent that executed the vocal action in the Observed Human condition (i.e., the experimenter). Because the extent to which agents are embodied is an important aspect of action perception and intention attribution (Hostetter & Alibali, 2008; Niedenthal et al., 2005), we suggest that future research should investigate whether temporal binding is produced when observing actions carried out by embodied agents other than human actors, or if in fact it is an exclusive feature of observed human actions.

Taken together, these findings suggest that audio-visuomotor information plays a role in the generation of the temporal binding effect over observed actions. Still, this hypothesis remains speculative based on preliminary findings. Future research should directly focus on this topic, with the scope to understand specific contributions of visual information and accessibility in observed actions.

The Role of Agency in Interaction

Our findings are also consistent with Kunde and colleagues (2017), who proposed a theoretical framework of sociomotor action control. According to this model, others' responses to our communicative actions (including our vocalizations) are used to plan subsequent communicative actions. Hence, being able to detect whether our actions were effective in producing the desired outcome acquires critical importance. For example, when a friend does not respond when we call their name, it may mean that our intention (to draw their attention) has not been achieved, and further actions are needed to achieve our goal. Indeed, experiencing agency over the outcomes we produce with our communicative actions is necessary for planning what to do next. Thus, the role of agency in social interaction is of central importance, and may also support higher-level mechanisms such as theory of mind. In these terms, agency may be a critical link in the chain connecting joint action and social cooperation, as ascribing agency to others (i.e., to perceive them as agentic, capable of producing changes in the environment) is crucial to develop expectations about their intentions and mental states (Sebanz & Knoblich, 2009).

Another possible limitation to the current research relates to the nature of the outcome that followed the action. While our work was motivated by the goal of shedding light on the social aspect of the action (i.e., talking), that action was not followed by a meaningful social outcome. This was essential to isolate the phenomenon of temporal binding for vocal actions. However future research might fruitfully explore how the relationship between communicative actions and their outcomes (e.g., by matching the outcome to the vocal command) impacts the experience of agency.

These findings may be of direct interest for developers of vocal assistants like Siri and Alexa - those user-friendly devices of increasing prevalence worldwide. During the design process, they should take into account that both robotic and human voices may not be perceived as agentic per se. This should direct future research into exploring different interfaces that allow more intuitive and spontaneous interactions, focusing on other typical factors of human-human interaction such as embodiment (Niedenthal et al., 2005) and temporal coordination (Schmidt et al., 2011). New insights and innovation in the development of vocal artificial agents will enhance the quality of the social engagement we experience towards them, which is likely to be a central element of the social interactions we will build in the near future.

Conclusions

Here we have reported consistent evidence of a novel temporal binding effect for vocal actions that produce a systematic outcome. Such effects occurred not only when participants produced the vocal action themselves, but also when they observed someone else doing it, as long as direct visual access to the other’s vocal action was possible. These findings make an important contribution to the growing literature concerned with how perceivers represent intentional actions and their consequences, demonstrating that action is not limited to movements by our hands and feet, but include a range of social behaviours as well.

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109.

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, *51*(8), 1619-1629.

Beyer, F., Sidarus, N., Fleming, S., & Haggard, P. (2018). Losing control in social situations: how the presence of others affects neural processes related to sense of agency. *eNeuro*, *5*(1), ENEURO-0336.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in cognitive sciences*, *6*(6), 237-242.

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., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological Science*, *20*(10), 1221-1228.

Capozzi, F., Becchio, C., Garbarini, F., Savazzi, S., & Pia, L. (2016). Temporal perception in joint action: This is MY action. *Consciousness and Cognition*, *40*, 26-33.

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current Biology*, *26*(5), 585-592.

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(1), 1-8.

Cravo, A. M., Claessens, P. M., & Baldo, M. V. (2011). The relation between action, predictability and temporal contiguity in temporal binding. *Acta Psychologica*, *136*(1), 157-166.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176-180.

Engbert, K., & Wohlschläger, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition*, *16*(2), 255-264.

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.

Engbert, K., Wohlschläger, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1261.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*(5722), 662-667.

Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, *30*(1), 15-31.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390-R392.

Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition*, *12*(4), 695-707.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243-267.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514.

Howard, E. E., Edwards, S. G., & Bayliss, A. P. (2016). Physical and mental effort disrupts the implicit sense of agency. *Cognition*, *157*, 114-125.

Humphreys, G. R., & Buehner, M. J. (2010). Temporal binding of action and effect in interval reproduction. *Experimental Brain Research*, *203*(2), 465-470.

Kunde, W., Weller, L., & Pfister, R. (2017). Sociomotor action control. *Psychonomic Bulletin & Review*, *25*(3), 917-931.

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.

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.

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, *18*(4), 1056-1064.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Consciousness and Cognition*, *21(1)*, 546-561.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*(3), 184-211.

Obhi, S. S., & Hall, P. (2011a). Sense of agency and intentional binding in joint action. *Experimental Brain Research, 211*(3-4), 655.

Obhi, S. S., & Hall, P. (2011b). Sense of agency in joint action: Influence of human and computer co-actors. *Experimental Brain Research*, *211*(3-4), 663-670.

Pfeiffer, U., Schilbach, L., Timmermans, B., Jording, M., Bente, G., & Vogeley, K. (2012). Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Frontiers in Psychology*, *3*, 537.

Poonian, S. K., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. *Experimental Brain Research*, *229*(3), 419-427.

Pfister, R., Obhi, S. S., Rieger, M., & Wenke, D. (2014). Action and perception in social contexts: intentional binding for social action effects. *Frontiers in Human Neuroscience*, *8*, 667.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, *1*(4), 515-526.

Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*(2), 129-140.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169-192.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, *31*(6), 889-901.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353-367.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, *30*(5), 834-845.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, *25*(1), 76-89.

Stephenson, L. J., Edwards, S. G., Howard, E. E., & Bayliss, A. P. (2018). Eyes that bind us: Gaze leading induces an implicit sense of agency. *Cognition*, *172*, 124-133.

Strother, L., House, K. A., & Obhi, S. S. (2010). Subjective agency and awareness of shared actions. *Consciousness and Cognition*, *19*(1), 12-20.

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in Psychology*, *4*, 127.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: experiencing control over the movements of others. *Journal of Personality and Social Psychology*, *86*(6), 838.

Wohlschläger, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self-and other-generated actions. *Psychological Science*, *14*(6), 586-591.

Yabe, Y., Goodale, M. A., & MacDonald, P. A. (2019). Investigating the perceived timing of sensory events triggering actions in patients with Parkinson's disease and the effects of dopaminergic therapy. *Cortex*, *115*, 309-323.

Author Note

Luca Pascolini, Lisa J. Stephenson, Andrew P. Bayliss, Natalie A. Wyer, School of Psychology, University of East Anglia.

The authors obtained permission to reproduce or adapt any copyrighted material from other sources. We would like to thank S. Gareth Edwards for critical insights during early stages of this research. The research was supported by University of East Anglia Studentships to L. Pascolini and L. J. Stephenson and by a Leverhulme Trust Project RPG-2016-173 to APB.

Correspondence regarding this article may be addressed to Luca Pascolini, School of Psychology, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom, email: L.Pascolini@uea.ac.uk

Author Contributions

LP, AB, and NW conceived the study. All authors designed the experiments. LP created the materials, collected and analysed the data. LP, AB, and NW interpreted the data. LP drafted the manuscript. All authors provided critical revisions and approved the final version of the manuscript for submission.

Data Availability

Raw data and experimental stimuli used in this study has been made public to comply with good open science procedure. All files can be accessed at https://osf.io/jp57r/. This study was not preregistered.

Conflicts of Interest

Authors declare no conflict of interest for this project.

Tables

Table 1

Exp 1	**Control**	**Operant Manual**	**Operant Vocal**	**Observed Manual**
	1.05 (0.26)[a]	0.92 (0.19)[b]	0.96 (0.21)[b]	0.92 (0.20)[b]
Exp 2	**Control**	**Operant Vocal**	**Heard Vocal**	**Heard Artificial**
	0.95 (0.19)[a]	0.82 (0.19)[b]	0.92 (0.14)[a]	0.97 (0.20)[a]
Exp 3	**Control**	**Operant Vocal**	**Observed Vocal**	**Observed Artificial**
	0.91 (0.24)[a]	0.81 (0.27)[b]	0.86 (0.24)[b]	0.92 (0.24)[a]

Table 1. Mean Proportional Interval Reproductions for each condition across all experiments. Respective standard deviations are reported in brackets. Different superscripts within the same row indicate a significant difference between conditions ($p < 0.05$).

Figure Captions

Figure 1. Trial Procedure for the four conditions in Experiment 1.

Figure 2. Mean proportional interval reproduction across the four conditions in Experiment 1 (A), 2 (B), and 3 (C). Error bars represent the standard error of the mean. Significant differences from the Control condition are highlighted for the 0.05 (*) and 0.01 (**) level.

Figure 3. Trial Procedure for the four conditions in Experiment 2.

Figure 4. Trial Procedure for the four conditions in Experiment 3.

Figures

Figure 1

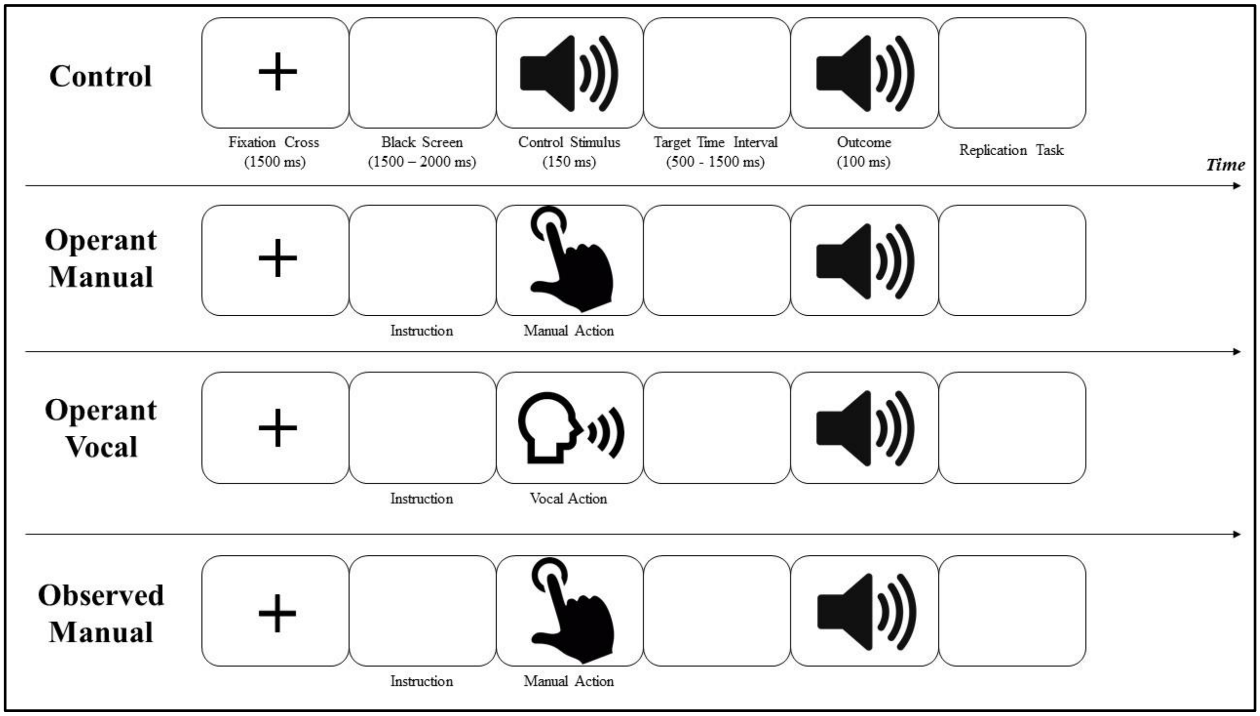


Figure 2

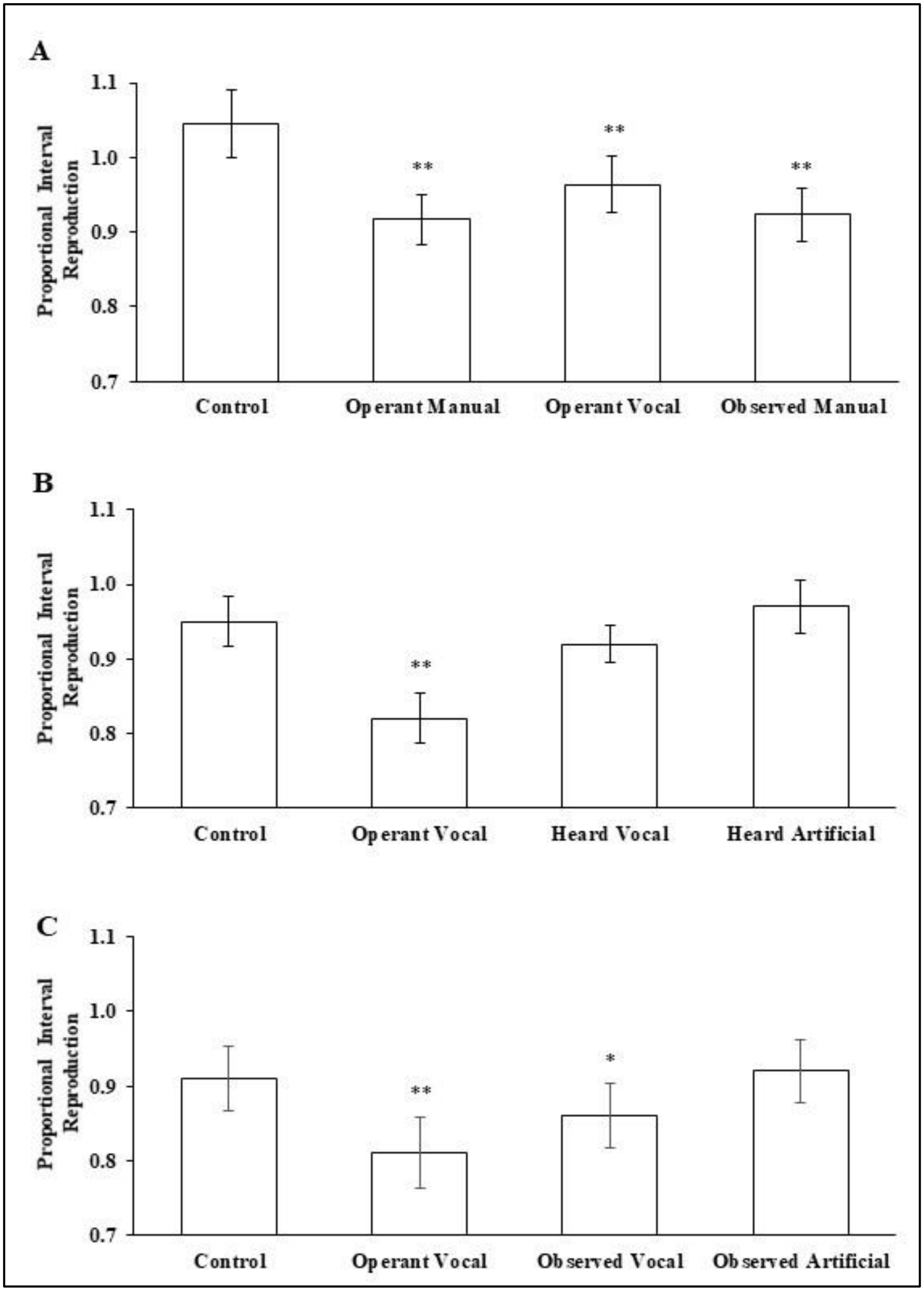

A
Proportional Interval Reproduction
1.1
1.0
0.9
0.8
0.7
**
**
**
Control
Operant Manual
Operant Vocal
Observed Manual
B
Proportional Interval Reproduction
1.1
1.0
0.9
0.8
0.7
**
Control
Operant Vocal
Heard Vocal
Heard Artificial
C
Proportional Interval Reproduction
1.1
1.0
0.9
0.8
0.7
**
*
Control
Operant Vocal
Observed Vocal
Observed Artificial

Figure 3

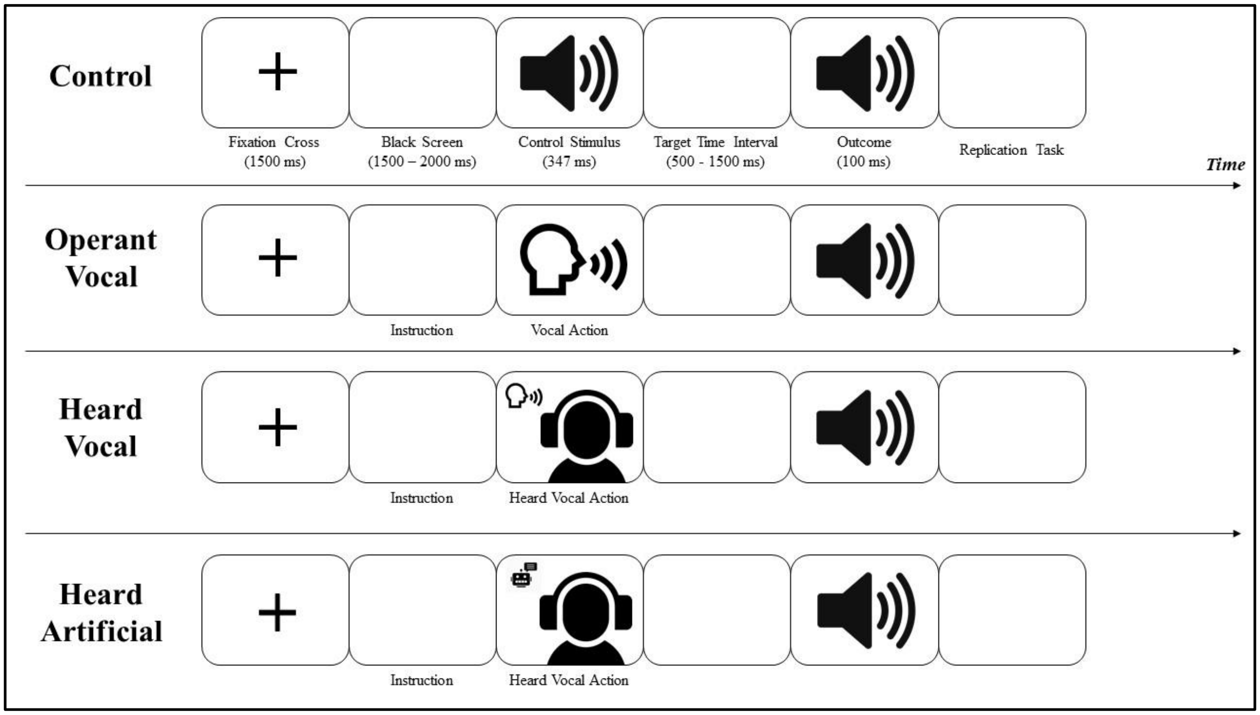
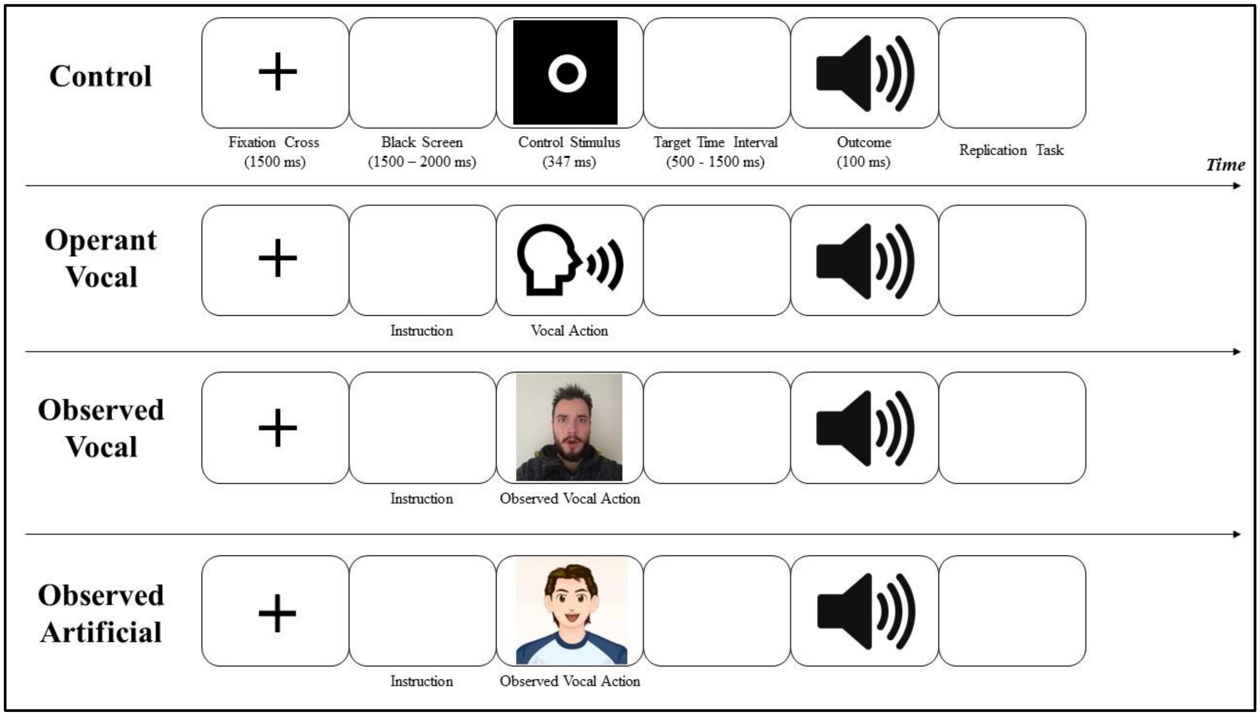

Control
Fixation Cross (1500 ms)
Black Screen (1500 – 2000 ms)
Control Stimulus (347 ms)
Target Time Interval (500 - 1500 ms)
Outcome (100 ms)
Replication Task
Time
Operant Vocal
Instruction
Vocal Action
Heard Vocal
Instruction
Heard Vocal Action
Heard Artificial
Instruction
Heard Vocal Action

Figure 4


Control
Fixation Cross (1500 ms)
Black Screen (1500 – 2000 ms)
Control Stimulus (347 ms)
Target Time Interval (500 - 1500 ms)
Outcome (100 ms)
Replication Task
Time
Operant Vocal
Instruction
Vocal Action
Observed Vocal
Instruction
Observed Vocal Action
Observed Artificial
Instruction
Observed Vocal Action