NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF PHILOSOPHY AND HISTORY OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS
DEPARTMENT OF PSYCHOLOGY
DEPARTMENT OF PHILOLOGY

Pinelopi-Panagiota Bounia-Mastrogianni
A.M. 15M10

Intentional binding as a function of action-effect interval and semantic relatedness

Thesis submitted
In partial fulfillment of the requirements for the Masters degree by the
Interdepartmental Graduate Program in Cognitive Science

Advisory committee:
Konstantinos Moutoussis, Associate Professor, National and Kapodistrian University of Athens
Argiro Vatakis, Assistant Professor, Panteion University of Social and Political Sciences
Nikolaos Smirnis, Associate Professor, National and Kapodistrian University of Athens

Athens, Greece
September 2018

The thesis is approved

Konstantinos Moutoussis...

Argiro Vatakis...

Nikolaos Smirnis..

© 2018 Pinelopi-Panagiota Bounia-Mastrogianni
Available online under a Creative Commons Attribution 4.0 Licence (International)
(CC-BY 4.0) https://creativecommons.org/licenses/by/4.0/

Abstract

Voluntary actions and their sensory effects are perceived closer in time; a phenomenon known as intentional binding (IB). Most up-to-date studies have examined IB employing one-modality action effects, mostly abstract, yet everyday life actions produce multisensory, informationally rich effects. Recently, Thanopoulos, Psarou, and Vatakis (2018) used naturalistic multisensory stimuli as action outcomes and showed that IB occurs when voluntary actions and their effects hold an inherent causal link from everyday experience. Given the short action-effect interval used in Thanopoulos & Vatakis's study (250ms; as in the majority of IB studies), in our first experiment, we manipulated this interval in order to investigate the limits of maintenance of IB in causal multisensory events. Using the same naturalistic stimuli and the same setup with Thanopoulos and Vatakis, we tested the participants in conditions varying in action intentionality and temporal predictability of the effect for intervals of 250, 800, 1000, and 1250ms using a simultaneity judgment (SJ) task. Further, given the use of a multisensory effect, the induction of IB may be affected by potential crossmodal binding rivalries. Particularly, the unity assumption might cause temporal stimulus shifts in order to reinforce a unified percept, possibly interacting with the temporal shift towards the action, as predicted by IB. Thus, in our second experiment, we investigated how strongly unified multisensory action effects can affect the IB phenomenon, using the same causal sequence of events and procedure as in Experiment 1 and varying the semantic content of the presentations. In our first experiment, the audiovisual pair was perceived earlier regardless of the presence of a voluntary action for intervals up to 1000ms, with the shifts becoming larger as the interval increased, revealing strong temporal binding. In our second experiment, only visual effects were perceived earlier, regardless of their congruency with the auditory stimulus or the presence of a voluntary action. Both results underline the strong influence of causal relations between our stimuli, that were necesssary for the temporal shifts to occur and clearly overcame the effects of intention.

Aknowledgements

I would like to express my sincere thanks to my co-supervisors, Prof. Konstantinos Moutoussis and Prof. Nikolaos Smyrnis for their guidance and assistance. My special gratitude to my supervisor, Prof. Argiro Vatakis, for her constant motivation, patience and support. Her extensive knowledge and perseverance made this work possible and set an inspiring example of a scientist for me as a young researcher.

I would also like to thank Vassilis Thanopoulos for his practical help and direct problem-solving throughout the process of this thesis.

Last, a big thanks to my friends and family, for their limitless ancouragement, support and understanding. Finally, I would like to thank everyone who participated in this study for their contribution.

Contents

List of figures

Intentional binding as a function of action-effect interval and semantic relatedness

1. Introduction

The experience of causing an event with our own actions holds a special place among the large number of factors that affect our subjective sense of timing. Humans tend to perceive their voluntary actions and their sensory effects as closer in time, a phenomenon referred to as intentional binding (IB; e.g., Haggard, Clark, & Kalogeras, 2002). As initially described, this temporal compression is experienced by subjectively shifting the percept of the action and its effect towards each other, only in cases when the action is self-generated (i.e., intentional; Haggard et al., 2002). Numerous studies, based on different research traditions, have focused on particular aspects and stages of the phenomenon such as action selection (e.g., Vastano, Pozzo, & Brass, 2017), sensory recalibration (e.g., Stetson, Cui, Montague, & Eagleman, 2006), and action-outcome congruency (e.g., Barlas & Kopp, 2018), as well as other closely related topics such as action observation (e.g., Wohlslager, Haggard, Gesierich, & Prinz, 2003). Apart from the interest on IB itself, IB paradigms have been widely utilised as an implicit measure for the sense of agency, the subjective feeling that we are the authors of our actions (SoA; Moore, Wegner, & Haggard, 2009).

It has been suggested that the experience of agency results from an optimal integration of both low level, sensorimotor cues (strongly influenced by motoric signals associated with voluntary action preparation and prediction; Blakemore, Wolpert & Frith, 2002), as well as of high level cues, based on extrinsic factors and inferential processes (Moore & Fletcher, 2012; Synofzik, Vosgerau, & Newen, 2008). This has led to the investigation of agency on different levels and to a consequent utilization of both explicit and implicit methods to better approach each factor that might contribute in the produced feeling (Moore & Fletcher, 2012). The former methods usually include the explicit attribution of an action, verbally or in a similar manner, to a certain agent (self or other), thus they better suit to the examination of second-level, inferential aspects of agency (Haggard, 2017). On the other hand, implicit measurements, and especially the utilization of the IB phenomenon, has saved researchers from the cognitive biases that stem from explicit judgments (e.g., people's tendency to overestimate their own agency over others'), while facilitating the examination of both first and second-level cues that construct the experience (Haggard, 2017). As already mentioned, differentiation on the IB methodologies has enabled the focus on the factors of intentionality, action selection and fluency, which are more evident on the early stages of action execution, as well as on the factor of causality, which is mostly based on the contiguity and contingency between the action and its produced effect (Haggard, 2017).

The special effect of intention on temporal perception has been emphasized on the preliminary studies that described IB, where the phenomenon was obtained only on the presence of a voluntary action, and disappeared when the action was involuntary (Tsakiris & Haggard, 2003) or when no action was involved (Haggard et al., 2002). Voluntary movements have been associated with distinctive neural events, such as preparatory activity in cognitive motor areas (e.g., the readiness potential, a characteristic slow negative electroencephalographic potential that occurs before movement, has traditionally been considered a marker of volition; Shibashaki & Hallett, 2006). Such neural activity has been reported to reflect stronger IB (Jo, Wittmann, Hinterberger, & Schmidt, 2014). However, mere temporal coincidence of TMS-induced involuntary movement and preparation of a voluntary action has not been sufficient to produce IB (Haggard & Clark, 2003), suggesting that cognitive preparation must precisely precede action execution. Similarly, Jensen, Vagnoni, Overgaard, and Haggard (2014) demonstrated that the mere presence of body movement, externally generated and, thus, lacking volition, is not enough for IB. On their study, participants had to recall previously executed voluntary or involuntary (TMS-induced) hand movements, made either with the same or opposite hands (Jensen et al., 2014). On the incongruent condition, when voluntary actions interfered with involuntary ones, people relied more on their intention to move than on their actual body movement to bring into memory the hand they had actually used during the trial, underlining the critical way in which intention shapes experience (Jensen et al., 2014). Interestingly, Desantis, Roussel, and Wazsak (2011) conducted a study, on which they manipulated the panticipants' beliefs on whether it was them or the experimenter that produced a tone. Although the tone was actually triggered by the participants' action at all times, the study showed that the mere belief that they were the agents on certain trials resulted on stronger IB on these trials than on the ones that where the agency was attributed to a third person (Desantis et al., 2011). These studies underscore that the manipulation of the intention of people's actions, regardless if that occurs on the neural or contextual level, directly influences the IB phenomenon.

While intentionality has been widely considered the determinant factor for IB, the strong causal relation between a goal-directed movement and its produced effects has also received a large amount of interest, such that it has even been suggested that the phenomenon should best be referred to as “causal binding” (Buehner & Humphreys, 2009). These approaches do not focus on the motoric aspect of IB, but rather consider it a special case of a more general causal binding, in a sense that human time perception, which is inherently noisy and ambiguous, often shifts closely presented events towards one another in time, especially if this paired presentation is frequent and agrees with prior knowledge (Buehner, 2012). Traditionally, causality has been considered an inference based on observable cues and not directly experienced (Hume, 1920), and such is the case of the relation between actions and their produced effects (Buehner & Humphreys, 2009; Wegner & Wheatley 1999). On this particular case, the causal relation is mainly inferred from three important factors: temporal contiguity, temporal predictability and contingency between actions and their effects (Moore & Obhi, 2012).

Various studies have attempted to access the action-effect causal relation examining its temporal parameters, namely the effect’s delay and temporal predictability. In Haggard et al. (2002) study, where researchers used a rotating clock method (Libet, Gleason, Wright, & Pearl, 1983), IB monotonously decreased as the action-effect interval increased for intervals between 250 and 650ms. Other studies have successfully induced IB for small intervals of 200-300ms, utilising both rotating clock and interval estimation methods, though not extending the effect delay interval beyond these timings (e.g., Engbert & Wohlschlager, 2007; Engbert, Wohlchlager, Thomas, & Haggard, 2007). However, studies that used interval estimation methods to measure the effect for longer delays, have reported IB for intervals up to 4s (Buehner & Humphreys, 2009, 2010; Nolden, Haering, & Kiesel, 2014). A recent study by Ruess, Thomaschke, and Kiesel (2017) investigated the time course of IB for time intervals between 100 and 400ms. Employing a rotating clock method as in Haggard et al. (2002) study, the researchers showed a peak in the temporal shift of the effect towards the action at 250ms (i.e., the shift was larger than on the 100 or 400ms delay), suggesting that the effect does not necessarily decrease in a monotonous manner with a greater action-effect delay (Ruess et al., 2017). Intriguingly, this was not the case with the temporal shift of the action towards the effect, that steadily increased as the effect was pulled further in time (Ruess et al., 2017), a result that contradicts the decrease that Haggard et al. (2002) had reported. Lastly, studies have used simultaneity judgment tasks (for SJ tasks in general see Vatakis, Navarra, Soto-Faraco, & Spence, 2008 and Zampini, Guest, Shore, & Spence, 2005), accordingly adapted in order to examine how the effect delay affects IB. Wenke and Haggard (2009) showed that temporal discrimination between two shocks was more difficult when they appeared closer to the action (150ms) than after a longer delay. On Cravo, Claessens, and Baldo's (2011) study, participants voluntarily produced an action, which was followed by a tone and a temporally independent flash, and had to judge whether the tone and the flash appeared simultaneously. People judged the tone as simultaneous with the flash when the latter preceded (as IB predicts) only when the tone followed at a fixed interval of 250ms, while the effect disappeared for random intervals between 250-700ms (Cravo et al., 2011). Although most studies seem to support stronger IB for short delays, which decreases as the delay increases, no consensus about the exact time course of the phenomenon has been reached.

On the other hand, most researchers have reached convergent conclusions regarding the effect of temporal predictability of the action outcome on IB. On the aforementioned Cravo et al. (2011) study, effects produced strong IB only when they followed the action at fixed, predictable intervals. This finding was replicated in Thanopoulos, Psarou, and Vatakis (2018) study, who adapted Cravo at al.'s (2011) study methodology using naturalistic stimuli. Also, Haggard and colleagues (2002) showed that effects occurring systematically at a certain time point after the participants' action, thus allowing for a temporal prediction, produced stronger IB than effects that occurred at a random time point, which could not be predicted in time. Similar results can be found in a number of IB studies (e.g., Desantis, Hughes, & Wazsak, 2012).

Various researchers investigating IB have chosen to investigate the action-effect causal relation by manipulating their contingency, mostly by presenting unexpected effects at various probabilities (e.g., Desantis et al., 2012; Haering & Kiesel, 2014). Moore, Dickinson, and Fletcher (2011) used an outcome-blocking paradigm (commonly utilized to demonstrate the importance of surprise in associative learning; Dickinson, 1981) combined with a rotating clock method (Haggard et al., 2002) to examine the effect of unexpected action outcomes to the perceived temporal delay. After associating two tones with two different hand-presses, the authors presented expected and unexpected tones as action effects (i.e., presses of a button, while a clock rotated) and asked their participants to report the timing of the hand-press or the timing of the tone. The participants' estimation was less accurate for both the action and the tone exact timing on trials when the tone was expected to occur, reflecting shifts on the perceived timing of the action and the effect towards each other. On the contrary, a reduced effect was obtained for surprising outcomes, on which participants judged the actual time on which actions and tones occurred more accurately (Moore et al., 2011). However, Desantis, Hughes, and Wazsak (2012) argued that temporal predictability alone is sufficient for IB, regardless of the effect's particular identity. On their research they used a rotating clock method and compared three conditions: participants carried out key-presses that produced either a certain pitched tone or one of two tones (high and low-pitched) on equal probability, or the tones were externally produced. On either case, they had to judge the tone's onset time (Desantis et al., 2012). The researchers observed that the condition with the consistent action's effect did not differ from the one with two equiprobable effects, but both of these two active conditions showed significant temporal compression, thus IB, compared to the respective passive conditions. On a second experiment, they examined the temporal judgments when the participant's action produced a congruent and an incongruent tone with respect to an association built on an earlier phase, and noticed that the identity of the tone had no effect on the judgment (Desantis et al., 2012). Similarly, Haering and Kiesel (2014) attempted to address the same question using the method of constant stimuli. Their participants made voluntary left or right hand-presses, which produced certain visual stimuli (coloured squares), previously associated with each hand. The squares appeared validly (i.e., after the associated hand-press) on 80% of the trials and were reversed on the rest (Haering & Kiesel, 2014). Afterwards, participants had to judge whether the action-effect delay they experienced was shorter or longer in duration than a comparison tone presented right after the sequence. The researchers observed that valid and invalid trials did not differ in terms of delay duration estimation, since the intervals were judged as shorter on both cases (Haering & Kiesel, 2014). Other studies have also attempted to sort out the influence of action-effect contingency on IB, reaching contradicting conclusions (Barlas & Kopp, 2018; Desantis, Wazsak, Moutsopoulou, & Haggard, 2016; Haggard, Poonian, & Walsh, 2009) and as of yet no final resolution has been reached.

Interestingly, some studies have attempted to violate the causal law that effects cannot precede their actions in the flow of time (Desantis et al., 2016; Rohde & Ernst, 2013; Stetson, Cui, Montague, & Eagleman, 2006). These approaches support that the brain takes into account the temporal relations between actions and events in order to adjust and maintain the causal sequence, recalibrating the motor act and its sensory consequences when necessary (motor-recalibration hypothesis; Stetson et al., 2006). Although these studies have mainly focused on the effects of temporal expectancy on the action-effect recalibration, Desantis et al. (2016) have also examined the effect of the specific outcome's identity on the temporal adjustments. Specifically, they conducted three experiments, on which participants had to judge whether a visual stimulus (dots moving upward or downward on a screen) preceded or followed a voluntary key-press. The visual motion was either congruent or incongruent with the participants' action, as learned on an earlier association phase, and could actually precede or follow the action. On the first experiment the participants were informed that the dot motion could be initiated by their own movement or by the computer depending on the appearance of the stimuli - such explicit reference was spared on the following two experiments (Desantis et al., 2016). Results of all three experiments clearly showed that congruent outcomes were more likely to be judged as following the action regardless of their actual temporal occurrence, suggesting that maintaining the action-effect causal sequence is also dependent on an action's specific sensory consequence (Desantis et al., 2016).

The vast majority of IB studies makes use of abstract stimuli (e.g., Buehner & Humphreys, 2009; Cravo et al., 2011; Desantis et al, 2011), thus the formation of a strong causal relation between the action and the effect requires the use of extensive adaptation and association strategies. Adaptation strategies allow for achieving greater contingency, given that the use of abstract stimuli makes the action-effect link arbitrary (e.g.,, Buehner & Humphreys, 2009; Thanopoulos et al., 2018). On one of the few attempts for an ecologically valid setup, Ebert and Wegner (2010) have used naturalistic visual stimuli as action effects (images of everyday objects). The images moved closer or further away on a screen, controlled with a joystick by the participant. The acts of pushing or pulling the joystick hold bodily significance and inherently create certain expectations regarding the stimulus movement, without the need for adaptation strategies (i.e., pulling will cause an object to move closer), enabling the researchers to manipulate action-effect consistency and check its effect on binding (Ebert & Wegner, 2010). Indeed, consistent action-effect sequences resulted in delays judged by the participants as shorter than they actually were, using an interval estimation method (Ebert & Wegner, 2010). At a recent study, Barlas and Kopp (2018) also employed a naturalistic setup that took advantage of familiar, built contingencies between stimuli. They had participants press an arrow key with the direction of their choice on a keyboard and their keypress resulted on a congruent or incongruent direction of an arrow presented on the screen, at four different levels of congruency (Barlas & Kopp, 2018). In the consequent interval estimation task, people systematically judged the intervals in congruent conditions as shorter (Barlas & Kopp, 2018). More recently, Thanopoulos et al. (2018) also showed that adaptation could be spared, if actions and effects hold an inherent causal link. On their experiments, they used familiar naturalistic stimuli and observed the IB effect without establishing a causal link through adaptation. Specifically, an image of a hand over a wooden surface served as fixation and a voluntary hand-press by the participant produced the subsequent effect after a fixed short interval. The effect consisted of the sound of the hand hitting the surface. An image of the hand hitting the surface also appeared at different stimulus onset asynchronies (SOAs) with the tone and participants had to judge whether the pair appeared simultaneously or not (simultaneity judgment; SJ). Blocks where no voluntary action was required were also used, as well as blocks of random intervals between the action and the effect. Since the particular sequence of events was familiar, the stimuli were inherently linked and expected to "follow" each other. Results revealed an IB effect for blocks where the voluntary action was followed by the effect on a fixed delay. This was not the case when abstract or somehow causally unlinked events were utilized, despite the common experimental setup (Thanopoulos et al., 2018). These findings suggest that naturalistic action-effect sequences ensure larger ecological validity, making use of the pre-existing causal relations between everyday actions and their sensory consequences.

At this point, we need to stress that most procedures examining the IB, use unimodal effects (often auditory or visual; e.g., Buehner & Humphreys, 2009; Haering & Kiesel, 2015; Rodhe, Greiner, & Ernst, 2014). However, everyday life actions usually have multisensory effects, where different modalities probably interact, influencing the temporal dynamics of action-effect perception. For instance, Kemenade, Arikan, Kircher, and Straube (2016) used bimodal (visual and auditory) outcomes to self-generated actions, aiming to examine how predictions for action effects on different modalities influence the perceived duration of delays between action and effect. On their study, they first presented trials with visual and auditory stimuli (a dot and a tone) or trials with only one of these stimuli at different delays after the participants' action, and they had to answer whether there was a delay or not. On bimodal trials, the delay of interest was between the action and the stated task-relevant modality (e.g.,, between the action and the tone, while a dot also appeared at different delays after the action). On a second experiment, a passive condition was added, where the button was pulled down at random times by the computer, followed by the same stimuli as in the first experiment. Results revealed greater accuracy on duration judgments for active bimodal trials, especially when the task-irrelevant modality stimulus was time-contiguous with the action, and vanished on passive stimuli presentation (Kamenade et al., 2016). Although IB was not the particular study's interest, results indicate important interplay between the modalities serving as action effects, suggesting possible ways in which temporal perception is modified by their interaction with voluntary action (Kamenade et al., 2016).

Few studies that more specifically address the IB phenomenon have utilised mutisensory action effects, such as the previously described in Cravo et al.'s (2011) and Thanopoulos et al.'s (2018) study. Both of these studies have presented auditory stimuli as predictable action effects, while visual stimuli were also presented at different SOAs with the auditory ones, resulting in a multisensory experience after the voluntary action (Cravo et al., 2011; Thanopoulos et al., 2018). While a temporal shift was evident for the fixed auditory stimulus alone, none of these studies looked into the possible interactions between the different modalities and the way these interactions might have affected IB. Such interactions are always present when stimuli are presented on different modalities within a particular temporal window, and might result to a strong multimodal integration of these stimuli to a unified percept, to a weak binding between them or to a clear segregation (Shams & Beierholm, 2012). Which percept will prevail depends both on low-level factors (e.g.,, spatial and temporal co-occurrence of different stimuli; Welch, 1999) and on whether the observer will judge that the stimuli belong together (i.e., the unity assumption; Chen & Vroomen, 2013; Vatakis & Spence, 2007; Welch & Warren, 1980). The unity assumption itself can be influenced by low-level factors, in the sense that spatiotemporal parameters can strengthen phenomenal causality and, thus, make integration more or less possible (Guski & Troje, 2003). However, the assumption that stimuli captured through different senses have a common environmental cause depends more on top-down cognitive factors, such as the semantic congruency between them, based on previous experience acquired from everyday life (e.g., Chen & Spence, 2017). Such congruency can yet be potentially built through consistent joined presentation of stimuli on an experimental setup (e.g., Radeau & Bertelson, 1974). Regardless of the context, previous experiments have reported temporal mislocations (i.e., ventriloquisms) of stimuli of different modalities, affected by the observers' unity judgements (Chen & Vroomen, 2013; Vatakis & Spence, 2007). In particular, when multimodal stimuli that serve as action effects are presented within the temporal window of integration on IB experiments, we can reasonably assume that their semantic congruency affects the possibility of integration and, thus, might lead to temporal ventriloquisms to satisfy a unified percept. For instance, on IB experiments that use abstract bimodal stimuli as effects (Parsons, Novich & Eagleman, 2013; Kemenade et al., 2016), consistent presentation of flashes and beeps within the integration window possibly caused some level of crossmodal binding or multisensory integration, leading to temporal shifts of the stimuli towards each other and, thus, affecting the temporal dynamics of perception. Even more, Thanopoulos et al. (2018), who used familiar naturalistic audiovisual stimuli - ensuring strong semantic congruency - report temporal shifts of the predictable outcome towards the action (i.e., the auditory stimulus), but shifts on the visual modality due to possible multisensory integration cannot be excluded.

At the present study, therefore, we intended to investigate two unresolved issues regarding the IB phenomenon, the impact of the action-effect delay and the impact of the semantic congruency of the effect on IB. We conducted two experiments, where we presented naturalistic multisensory stimuli as action effects on an IB paradigm, varying their temporal parameters and their semantic content, while participants had to complete a simultaneity judgment task. On both experiments, our methodology was adapted by Thanopoulos et al. (2018) study, using similar experimental setup and stimuli.

On our first experiment, we examined four different fixed and random action-effect delays ranging from 250 to 1250ms. Those intervals where chosen to cover the range from the commonly reported interval of 250ms to an interval shortly after 1s, since we presumed that larger delays will not be able to maintain the causal link between actions and effects, which is more strongly inferred in closely presented events (e.g., Ruess et al, 2017). We hypothesized that IB will occur only at the short intervals (250 and 800ms), but it will likely be disrupted for longer intervals in the range of 1000 to 1250 ms, which exceed the usually experienced timing between a particular action and its effect in everyday life (Haering & Kiesel, 2015). On our second experiment, we investigated the interaction between time compression predicted by the IB hypothesis and potential temporal ventriloquisms induced by interactions between multisensory effects We, thus, presented either auditory or visual abstract and naturalistic stimuli (same as in Experiment 1) as effects of the voluntary action, paired at multiple stimulus onset asynchronies (SOAs) with stimuli on the other modality stream that were either congruent or incongruent semantically. We expected that congruent stimuli will attract each other in time due to integration, often acting against the temporal shift that IB predicts or resulting in the temporal shift of the multisensory event as a whole and causing greater difficulty in the SJ task, whereas incongruent stimuli were expected to be perceived as separate events, allowing for larger temporal shifts of the fixed effect alone towards the action and easier simultaneity judgments.

2. Methods

2.1. Experiment 1

2.1.1. Participants

One hundred seventeen people took part in this experiment (87 females; Mean age = 20.6 years), pseudorandomly assigned to one of four groups with different action-effect interval: 250ms: 34 people (20 females; mean age = 21.9 years), 800ms: 31 people (26 females; mean age = 20.4 years), 1000ms: 26 people (21 females; mean age = 20.4 years), 1250ms: 26 people (20 females; mean age = 19.7 years). Participants reported normal or corrected-to-normal vision and normal hearing. All subjects were naïve to the purpose of our experiment and were awarded with course credit for their participation. The total duration of the experiment was 40 minutes.

2.1.2. Apparatus

The experiment took place in a dark and quiet room. Programming of the experiment was performed using OpenSesame (Version 2.9.7; Mathot, Schreij & Theeuwes, 2012). The visual stimuli were presented on a CRT monitor (1600 x 1200 pixel resolution, 60 Hz refresh rate), while the auditory stimuli were presented through two loudspeakers placed on the right and left side of the monitor (Creative Inspire 265). The participants sat at a 60 cm distance from the monitor. Voluntary actions were carried out via the Griffin Technology PowerMate USB Multimedia Controller v.2.0.1.

2.1.3. Stimuli

The experimental stimuli were two images, one serving as fixation and one as pair to the auditory action effect, and an auditory tone. Specifically, the fixation point was an image of a static hand over a wooden surface (416 x 331 pixels; Figure 1A). The second visual stimulus consisted of the view of a static hand hitting a wooden surface (416 x 331 pixels; Figure 1B), while the auditory stimulus (Figure 1C) was the impact sound of the hand hitting the wooden surface (sampling frequency of 96000 Hz, stereo, 76 dB). That is, the fixation screen- static hand over wooden surface- matched with the SJ stimulus pair –auditory and visual impact of the hand on the wooden surface.

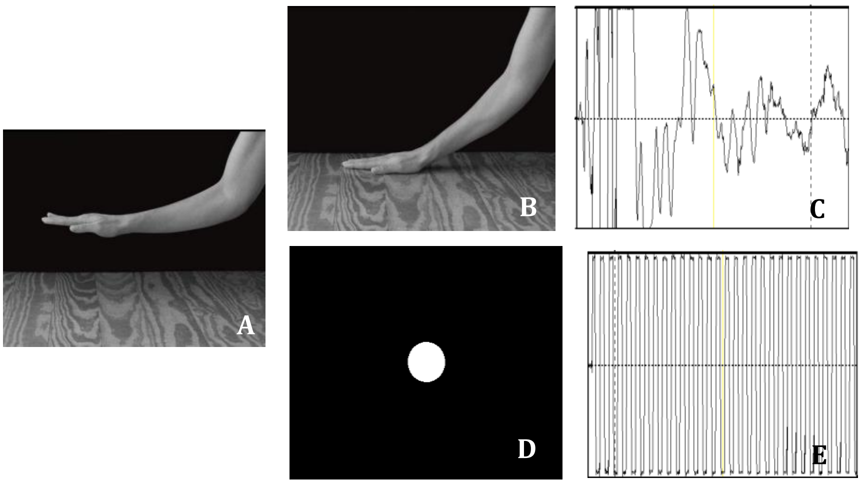


Figure 1. The visual and auditory stimuli used in the two experiments consisted of: A) a static image of a hand over a surface (used as fixation image on both experiments), B) a static image of a hand hitting a surface (used on both experiments), C) an impact sound of a hand hitting a wooden surface (used on both experiments), D) a white circle (used on Exp.2) and E) an auditory tone (used on Exp.2).

2.1.4. Design

The experiment consisted of "Action" and "No-Action" trials, depending on the presence of a voluntary action on the beginning of each trial (as in Thanopoulos et al.. 2018). On the "Action" conditions, the fixation image remained on screen until the action (i.e., pressing on the Powermate knob) was executed by the participants, at a time of their choice. On the "No-Action" conditions, the initial fixation image disappeared randomly after 1000-2000 ms. Following the hand-press or the fixation disappearance, the tone was presented after an fixed interval ("Fixed" conditions) or a random interval ("Random" conditions). On "Fixed" conditions, the tone was presented after 250, 800, 1000, or 1250ms. On "Random" conditions, the tone was presented at a random interval between 0 and 500ms over the corresponding fixed interval. On all conditions, the visual stimulus appeared at nine different onset asynchronies (SOAs) relative to the tone (SOAs: 0, ±60, 100, 150, and 200 ms; negative SOAs indicate the visual stimulus is presented first). Both stimuli remained on screen for 30ms. Afterwards, participants had to indicate whether the pair of stimuli was presented simultaneously or not, pressing one of two keys on a keyboard ("a" for asynchronous, or "s" for synchronous stimuli). Together, all these resulted to four different combined conditions ("Action-Fixed", "Action-Random", "No-Action-Fixed" and "No-Action - Random") for each of the four intervals of the experiment, thus to sixteen different experimental blocks. Each participant completed all four action and interval type conditions for one action-effect interval.

2.1.5. Procedure

At the beginning of the experimental procedure, all participants completed an adaptation block and a practice block. The adaptation block was similar to the experimental block (in terms of the action-effect interval of the following experimental block and the effect's interval type) but without the visual stimulation, and was inserted to form an association between the participant's action and its effect (i.e., the tone) on

"Action" trials. The same process was also employed on "No-Action" blocks, for a comparable procedure. The adaptation block had 5 trials and participants completed it before each of the four experimental blocks.

The practice block followed the adaptation block and was used to familiarize the participants with the experimental procedure. It was identical to the experimental block, only with fewer SOAs (i.e., 0 and ±200ms) and participants had to complete at least one block with 9 trials and answer correctly at least one trial with a certain SOA to continue to the experimental block. If they failed to answer, the practice block was repeated.

Next, the participants continued to the experimental session, in which they completed the main experimental blocks ("Action-Fixed", "Action-Random", "No-Action-Fixed" and "No-Action -Random"). They were asked to fixate on the initial image, which disappeared after a handpress or a certain time interval, depending on the block conditions, then the experimental stimulus pair was presented and they were asked to provide an answer regarding its simultaneity. Participants were informed that the task was unspeeded and they should answer confidently. The next trial followed only after an answer was given. The main experimental blocks were presented in a pseudo-randomized order. Stimulus presentation was also pseudo-randomized. Each experimental block consisted of 9 trials (i.e., 9 different SOAs) with 10 repetitions resulting in a total of 90 experimental trials (Figure 2).

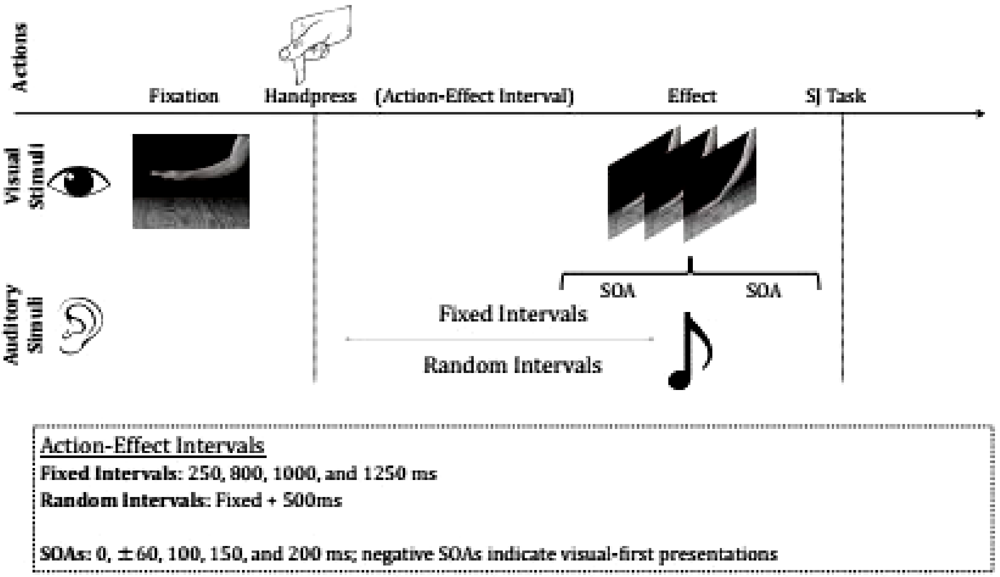


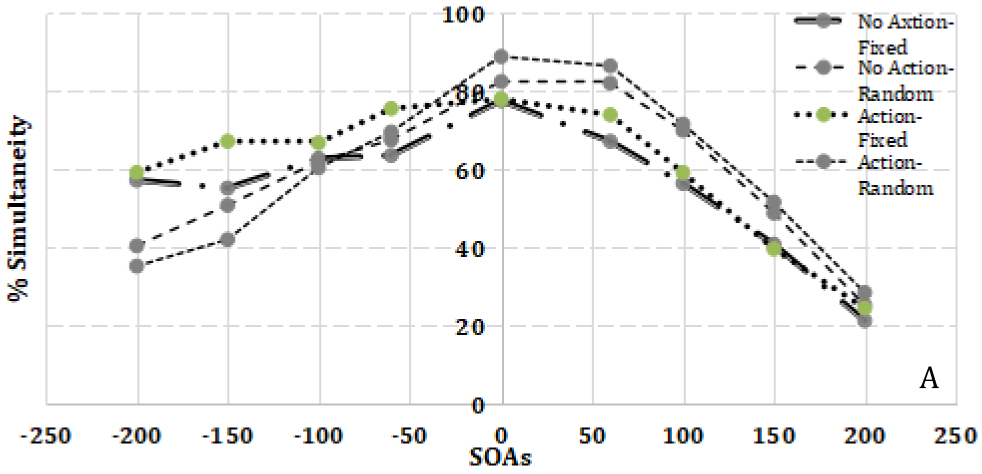
Figure 2. Schematic representation of the procedure followed on Exp.1

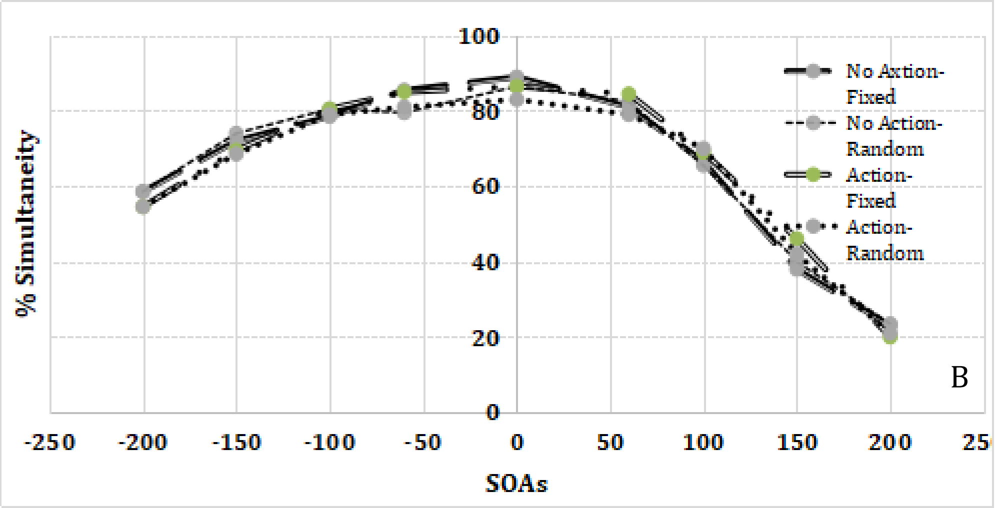
2.1.6. Results and Discussion

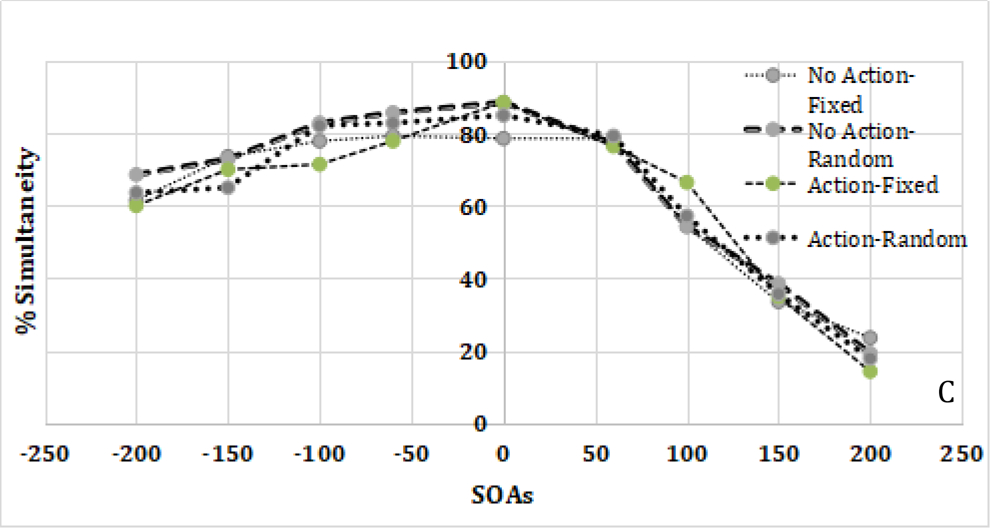
As already mentioned, we used a SJ task (e.g., Vatakis et al., 2008; Zampini et al., 2005), where participants had to judge whether the audiovisual events that served as action effects were simultaneous or successive. All participants' responses were averaged into a percentage of "simultaneous" responses and plotted as a function of SOA for all four experimental conditions for each interval (Figure 3A-D). Figure 3 shows that the participants were more likely to judge the pair of audiovisual stimuli as being simultaneous on the negative range of SOAs (i.e., when the visual stimuli were presented first) on all intervals except the longer one (1250ms). This suggests that participants generally experienced more difficulty in determining the asynchrony between the events when the visual stream was leading, while the Interval Type seems to have affected the SJs only for the smaller interval (250ms), where Random intervals seem to have been easier than Fixed ones when the visual stimuli preceded.

We fitted the observed distribution of responses to a Gaussian function for each participant, using the maximum likelihood estimation (see Myung, 2003). We compared the observed distribution with a normal distribution for each participant using the nonparametric Kolmogorov–Smirnov goodness-of-fit test (Vatakis et al., 2008; Zampini et al., 2005). This allowed us to obtain three parameters: the PSS (i.e., point of subjective simultaneity), which expresses the timing between two events at which a participant is most likely to give a synchronous response (a negative PSS value implies that the synchrony is perceived when a visual stimulus precedes an auditory stimulus, thus the auditory stimulus is shifted towards the action); the SD (i.e., standard deviation of the distribution), an estimate of the spread of the fitted distribution that provides a measure of difficulty for the temporal discrimination task across the SOA range tested (i.e., smaller SD values indicating better discrimination performance; Vatakis et al., 2008; Zampini et al., 2005); and the "peak of probability", which defines the peak height of the distribution for the SJ task with higher peaks indicating an increased probability of making a "simultaneous" response.

The PSS, SD, and A values were analyzed through a mixed analysis of variance (ANOVA) with the factors of Action (Action, No Action) and Interval Type (Fixed, Random) as within-subjects factors and Delay Interval as between-subjects factor. LSD tests were used for all post-hoc comparisons. For the PSS parameter, we







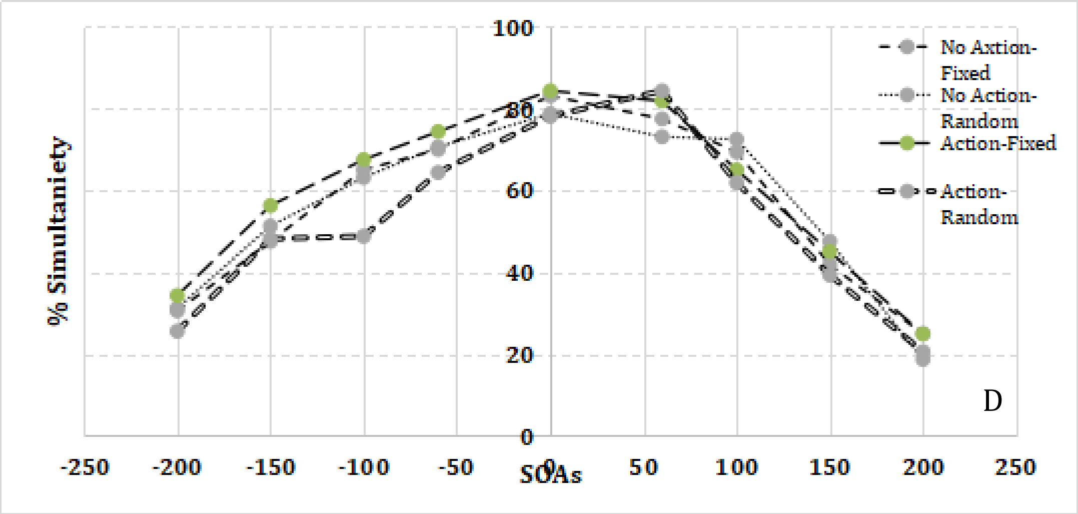


Figure 3. Percentage of “simultaneous” responses, plotted as a function of stimulus onset asyncronies (SOAs) between auditory and visual stimuli for all experimental conditions tested (No Action-Fixed. No Action-Random, Action-Fixed, Action-Random) for delays of A)250ms, B)800ms, C)1000ms and D)1250ms.

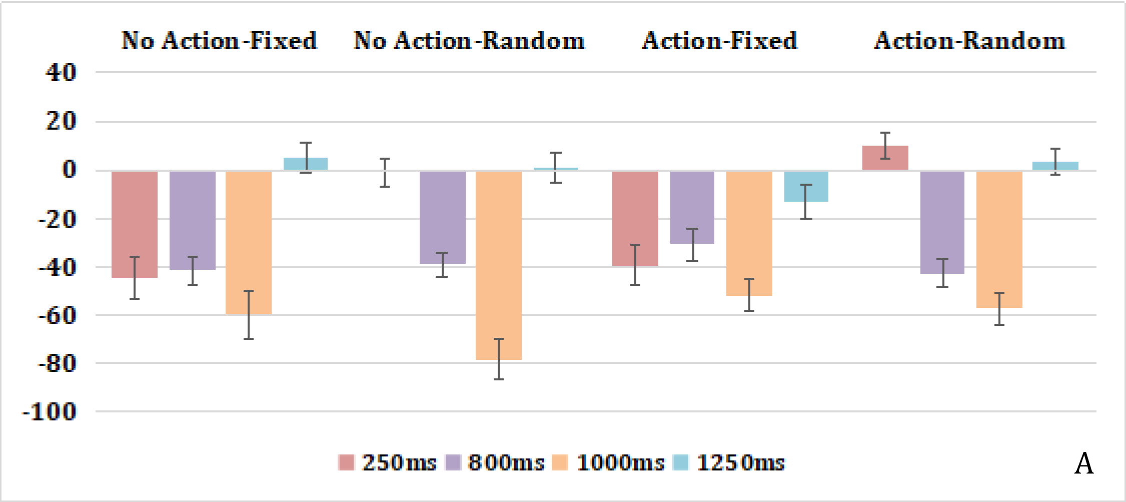
checked if its values were within range of SOAs tested. If some values were outside those boundaries, this meant that participants were unable to perform the task, thus we discarded their data and excluded them from further analysis.

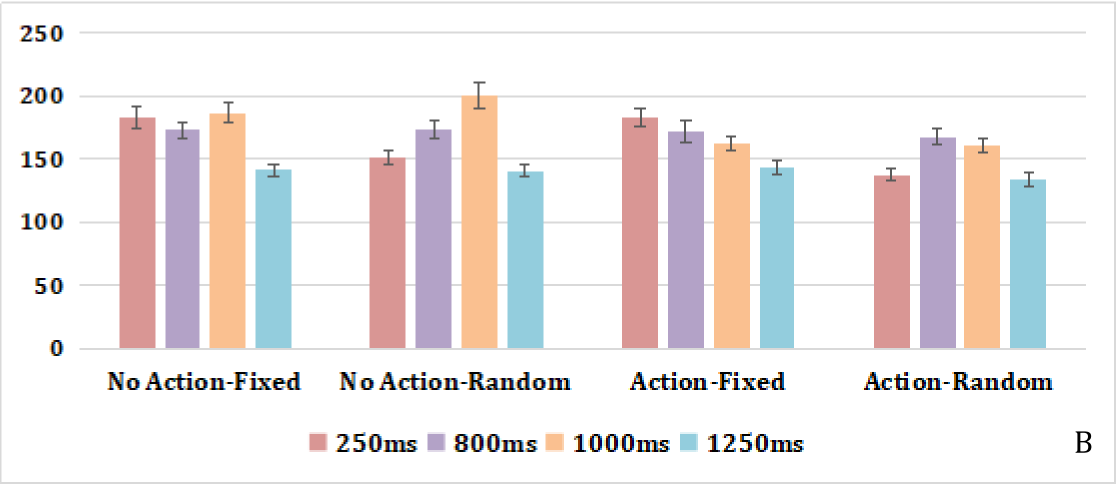
Forty-two participants were excluded from further analysis due to their PSS values (nine from 250ms interval, eleven from 800ms interval, twelve from 1000ms interval and ten from 1250ms interval). The analysis of the PSS values did not result in a significant main effect of Action [F(1, 71) = 0.078, p = 0.780, $\eta^2 = 0.0003$], but a significant main effect of Interval Type [F(1, 71) = 4.110, p = 0.046, $\eta^2 = 0.02$], with Fixed intervals being more negative (M = -31.92ms) than Random ones (M = -21.76ms), revealing temporal binding, regardless of the presence of a voluntary action. Such effect could suggest that, since the causal sequence of events is maintained, the temporally predictable effect is shifted towards its first event and perceived closer to it. A significant main effect of Delay Interval was also obtained [F(3, 71) = 3.031, p = 0.035, $\eta^2 = 0.114$], with participants experiencing significantly greater shifts of the tone towards the action on the 1000ms delay (M = -49.264ms) than on the 250ms delay (M = -18.630ms) and the 1250ms delay (M = -0.864ms), while the shifts on the 800ms delay were also significantly greater than on the 1250ms delay. Although seemingly contradicting previous research that shows decrease of the shifts as the action-effect delay increases, our results are likely explained by the causal link between the fixation image and the multisensory pair that follows. This link is probably maintained up to 1s, causing larger shifts as the interval itself is larger, but breaks for the biggest interval (1250ms). The interaction between Interval Type and Delay Interval was also significant [$F(3, 71) = 7.738$, $p < 0.001$, $\eta^2 = 0.213$], showing that, only for the 250ms delay, Fixed intervals had significantly more negative PSS values than Random Intervals (Mean Difference = 46.807ms). In other words, for the shortest action-effect interval, participants shifted the tone only when it was predictable, regardless of the presence of a voluntary action. Further, for Random intervals, on the 250ms and 1250ms delay, participants had significantly more positive PSS values (M = 4.773ms and M = 2.425ms respectively) than on the 800ms (M = -41.039ms) and the 1000ms delay (M = -53.204ms) (Figure 4A). This difference between the shifts in the 250ms and the shifts in the next two intervals may have occurred due to the fact that, for intervals closer to the beginning of the sequence, low-level temporal parameters such as predictability might play a more crucial role in determining causality than high-level factors such as semantics, which are probably more informative in greater intervals. .

The analysis of SD revealed that neither the main effect of Action [$F(1,71) = 1.417$, $p = 0.238$, $\eta2 = 0.007$], nor of Delay Interval [$F(3, 71) = 1.759$, $p = 0.163$, $\eta2 = 0.07$] were significant, but the main effect of Interval Type was [$F(1, 71) = 4.773$, $p =0.032$, $\eta^2 = 0.02$], showing that on Fixed intervals the temporal discrimination was significantly more difficult. The interaction between Interval Type and Delay Interval was also significant [$F(3, 71) = 4.913$, $p = 0.004$, $\eta2 = 0.06$]. Specifically, only on the 250ms delay, Fixed intervals had significantly greater SD values (Mean Difference = 38.585), implying a greater difficulty on these trials. This difficulty does not necessarily contradict the general tendency of participants to judge the tone as simultaneous with the image when the image preceded on fixed trials, but rather reveal possible integration of the multisensory pair to a unified percept, more difficult to segregate when the causal sequence was satisfied with temporally predictable events (see also General Discussion). Trials with fixed intervals were also significantly more difficult on the 250ms delay (M = 182.978) than on the 1250ms delay (M = 142.139), while trials with Random intervals were easier on 1250ms delay (M = 137.272) than on 800ms (M =

170.400) and 1000ms delay (M = 170.072) and on 250ms (M = 144.393) than 800ms delay.

Lastly, the analysis of the peak of probability revealed no significant main effects of Action [F(1, 71) = 2.527, p = 0.116, η^2 = 0.012], Interval Type [F(1, 71) = 2.081, p = 0.154, η^2 = 0.007], and Delay Interval [F(3, 71] = 0.689, p = 0.562, η^2 = 0.03]. However, the interaction of Interval Type and Delay Interval was significant [F(3, 71) = 5.927, p = 0.001, η^2 = 0.06], with participants on the 250ms delay being more probable to judge the audiovisual pair as simultaneous on Random interval trials than on Fixed ones (Mean Difference = 0.063; Figure 4B, C), This finding matches the previous results for the 250ms condition, further supporting that participants did not judge the audiovisual pair as simultaneous on Fixed conditions in general, but they obviously did on the particular case when the image preceded the tone, leading to high SD and shifted PSS values and probably, as already mentioned,





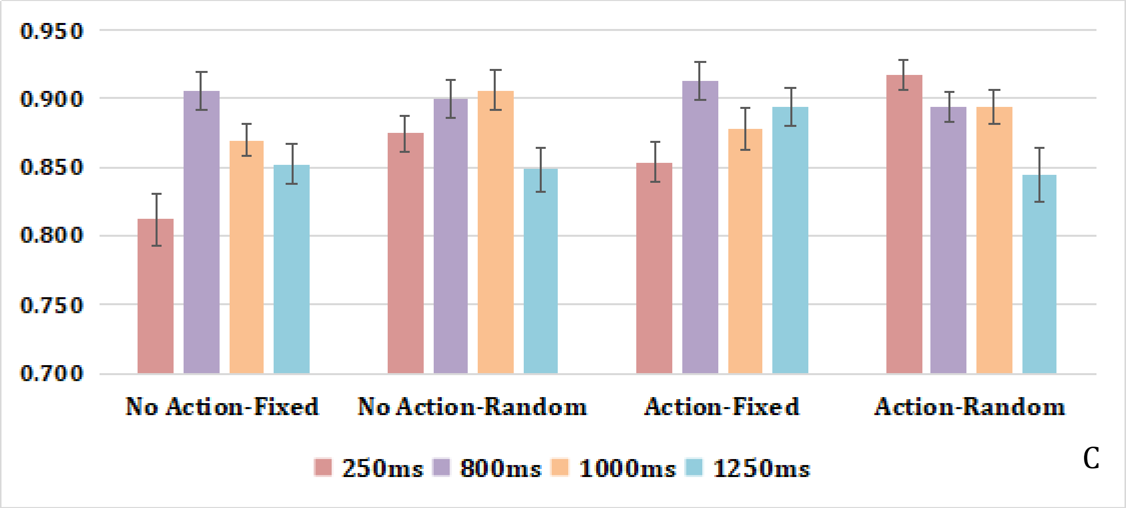


Figure 4. Mean values of A)PSS, B)SD and C)peak of probability for all experimental conditions tested, for all action-effect delays.

2.2. Experiment 2

2.2.1. Participants

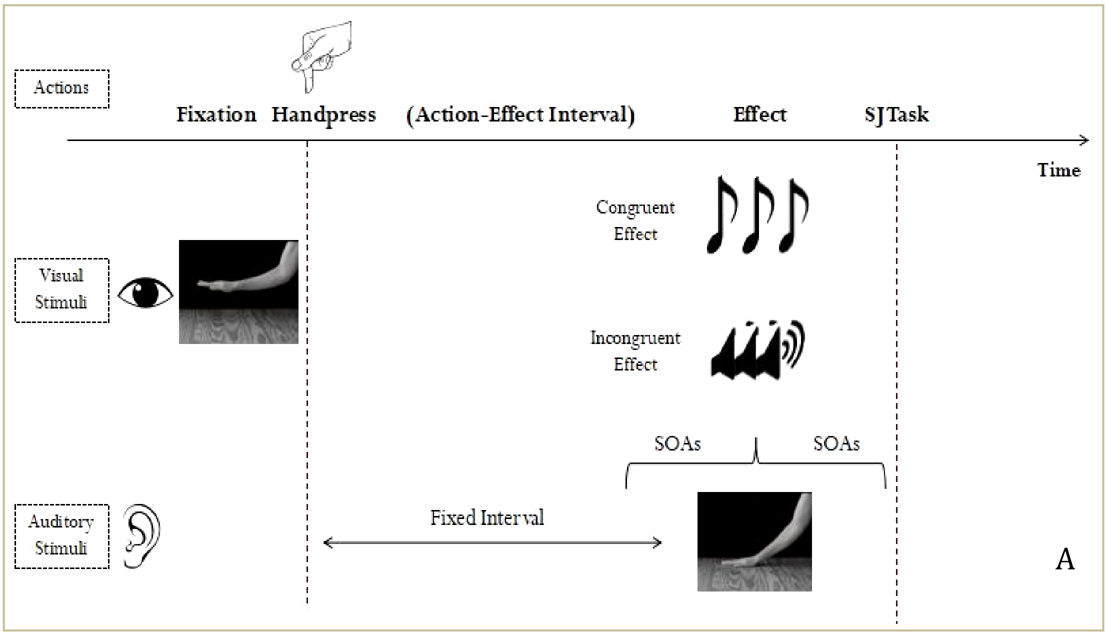
Fifty-nine new people participated in this experiment (47 females; mean age=18.8 years). They reported normal or corrected-to-normal vision and normal hearing. All subjects were naïve to the purpose of our experiment and students were awarded with course credit for their participation. The total duration of the experiment was 60 minutes.

2.2.2. Apparatus and stimuli

The same apparatus and setting as in Experiment 1 was used. We used the same visual stimulus as fixation, but two new abstract stimuli were added to the experimental pair. The abstract visual stimulus was a white circle (Figure 1C) with a 42 mm radius and was presented on the centre of the screen, while the tone (Figure 1B) had a sampling frequency of 44100Hz (stereo) and presented at 74 dB (Figure 1D). The naturalistic stimuli were the same as in Experiment 1.

2.2.3. Design

The experiment consisted of "Action" and "No-Action" trials, depending on the presence of a voluntary action at the beginning of each trial, similarly as in Experiment 1. Again, on the "Action" conditions, the fixation image remained on screen until the action was executed by the participant, while on the “No-Action” conditions, the initial fixation image disappeared randomly after 1000-2000 ms. On this experiment, following the hand-press or the fixation disappearance, the effect stimulus was always presented after a fixed interval of 250ms. The effect stimulus was either the naturalistic image or the auditory tone from Experiment 1, resulting on either “Visual” or “Auditory” conditions. This resulted to four experimental conditions (“Action-Visual”, “Action-Auditory”, “No-Action-Visual”, No-Action- Auditory”). Paired with each effect, a congruent or incongruent stimulus on the corresponding modality was presented. Congruent stimuli formed a naturalistic pair (image and impact sound of hand hitting surface), regardless of the effect modality, while incongruent stimuli resulted in a mismatching pair of a naturalistic and an abstract stimulus. Congruency varied from trial to trial, with equal probability of congruent and incongruent events. All stimuli remained on screen for 30ms. Based on the effect stimulus modality on a given block (e.g., “visual”), the other modality stimulus (e.g., auditory) appeared at nine different SOAs relative to the effect (SOAs: 0, ±60, 100, 150, and 200 ms; negative SOAs indicate the no-effect stimulus was presented first). Afterwards, participants had to answer whether the effect pair was presented simultaneously or not, pressing one of two keys on a keyboard (“a” for asynchronous, or “s” for synchronous stimuli).



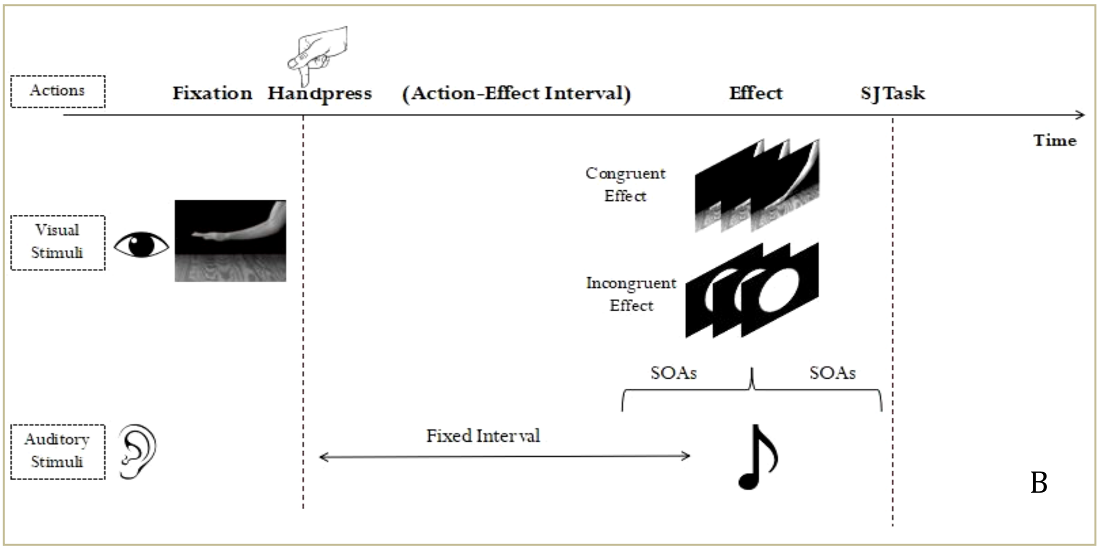


Figure 5. Schematic representation of the procedure followed on Exp.2, on A) blocks with a fixed visual effect and B) blocks with a fixed auditory effect

2.2.4. Procedure

The experiment followed the same procedure as in Experiment 1, with minor changes regarding the number of trials in the practice and experimental block: Participants completed 16 practice trials, in order to become familiarized with the possible outcome pairs, while the experimental session composed of 180 trials, due to the manipulation of congruency within each block (Figure 5A,B).

2.2.5. Results and Discussion

Thirty-three participants were excluded from further analysis due to inability to perform the task, as manifested in PSS values which fell outside the SOAs used in the experiment. The mean proportion of “simultaneous” responses showed that participants were more likely to judge the audiovisual pair as simultaneous on blocks with visual effects when the auditory stream was leading, regardless of the congruency of the stimuli (Figure 6A,B). This shows the negative temporal shift of the visual effect, while the exact opposite phenomenon is obvious for auditory effects: participants judged them as simultaneous when the visual stream followed the tone.

From the analysis of the PSS values, we obtained no significant main Effect of Action [F(1, 25) = 0.445, p = 0.511, η^2 = 0.001] and Congruency [F(1, 25) = 0.363, p =

0.552, η2 = 0.001]. However, a significant main effect of Effect Modality was revealed [F(1, 25) = 67.115, p < 0.001, η2 = 0.43], with the blocks

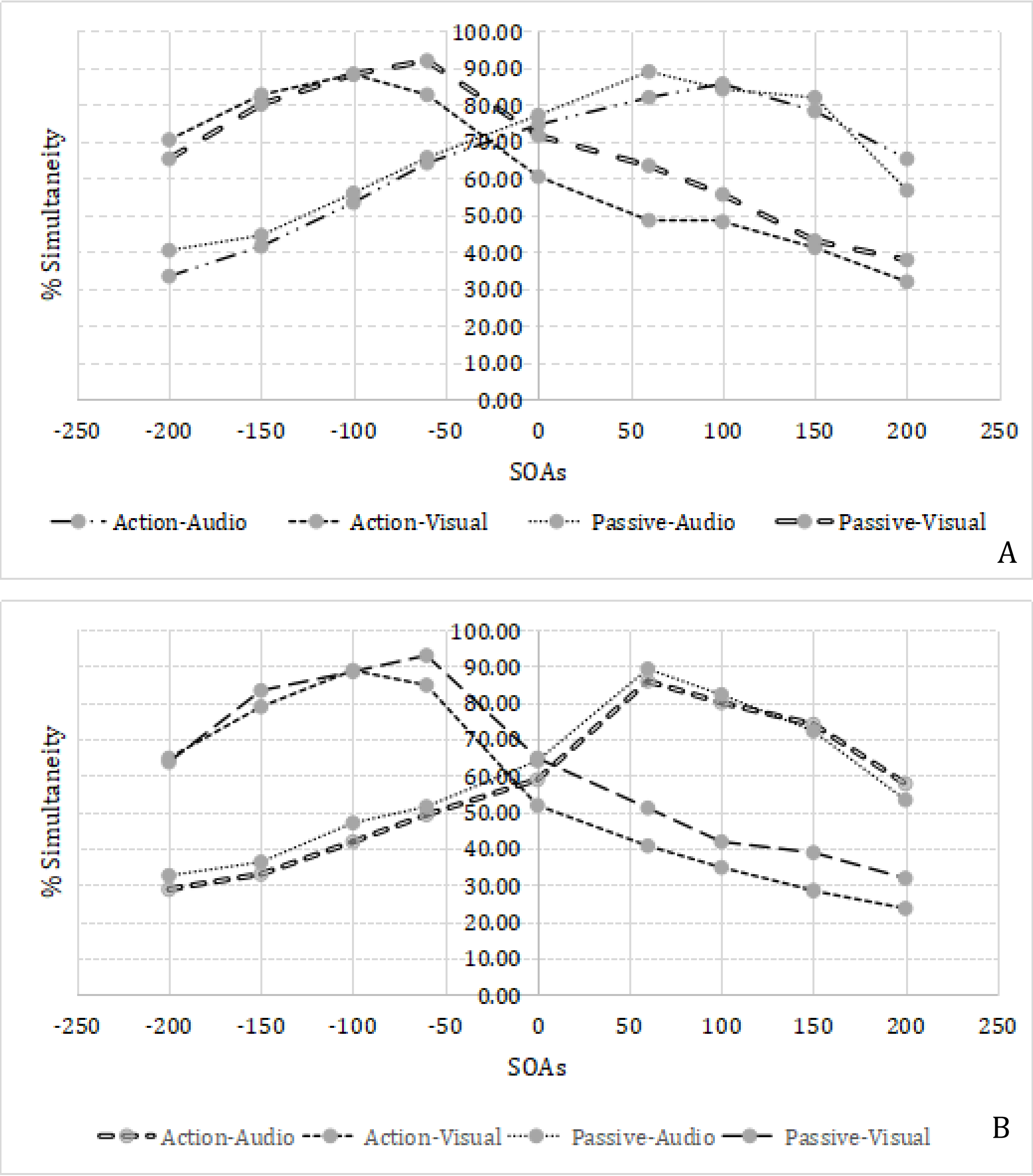


Figure 6. Percentage of “simultaneous” responses, plotted as a function of stimulus onset asyncronies (SOAs) between auditory and visual stimuli for all experimental conditions tested (Action-Audio, Action-Visual,Passive-Audio, Passive-Visual) for A)incongruent and B)congruent trials

with visual action effect having significantly lower PSS values (M = -106.656ms) than the ones with auditory effect (M = 73.243ms). This actually points out that only on blocks with a visual effect the participants perceived the effect as occurring earlier in time, while the opposite is the case with auditory effects. Further, analysis revealed a marginally significant interaction between Action and Effect Modality [F(1, 25) = 4.137, p = 0.053, η2 = 0.011]. Participants perceived the tone on auditory blocks as occurring later when they produced it with their voluntary action (Action block) than on the No-Action block (Mean Difference = 19.706ms; Figure 7A). This finding, auditory effects being judged as simultaneous with lagging visual stimuli, seems particularly unexpected, since it is generally accepted that people are more tolerant to auditory than visual delays. However, within a general framework that takes into account the causal relation formed between the fixation image and the subsequent audiovisual pair, it is possible that, while the image on visual trials was indicative of the sequence's end, the auditory stimulus did not play such a role, and thus was not shifted towards the beginning of the sequence, but rather its perception was delayed (see also General Discussion).

The SD analysis showed a main effect of Congruency [F(1 ,25) = 6.776, p = 0.015, η2 = 0.06], with incongruent trials having significantly greater SD values (M = 209.735) than congruent ones (M = 175.605). This difference suggests that participants found incongruent trials significantly more difficult than congruent ones. This is likely explained by the pairing of the complex stimulus effects with simpler abstract stimuli, which were probably perceived as shorter in duration, making the criterion for the SJ task more variable from trial to trial. On the contrary, no significant effect of Action [F(1, 25) = 0.210, p = 0.651, η2 = 0,002] and Effect Modality [F(1, 25) = 1.767, p = 0.196, η2 = 0,013] was found. Further, none of the interactions between Action and Modality [F(1, 25) = 0.291, p = 0.594, η2 = 0.002], Action and Congruency [F(1, 25) = 1.782, p = 0.194, η2 = 0.01] and Congruency and Modality [F(1, 25) = 0.097, p = 0.757, η2 = 0.001] were significant.

Lastly, the analysis of the peak of probability factor revealed no main effect of Congruency [F(1, 25) = 1.117, p = 0.301, η2 = 0.003], Effect Modality [F(1, 25) = 1.302, p = 0.265, η2 = 0.003] and Action [F(1, 25) = 2.092, p = 0.160, η2 = 0.006] (Figure 7B,C). No significant interaction between Action and Modality [F(1, 25) = 0.330, p = 0.571, η2 = 0.001], Action and Congruency [F(1, 25) = 0.562, p = 0.461, η2 = 0.001] and Congruency and Modality [F(1, 25) = 0.142, p = 0.709, η2 = 0.0004] was revealed either.

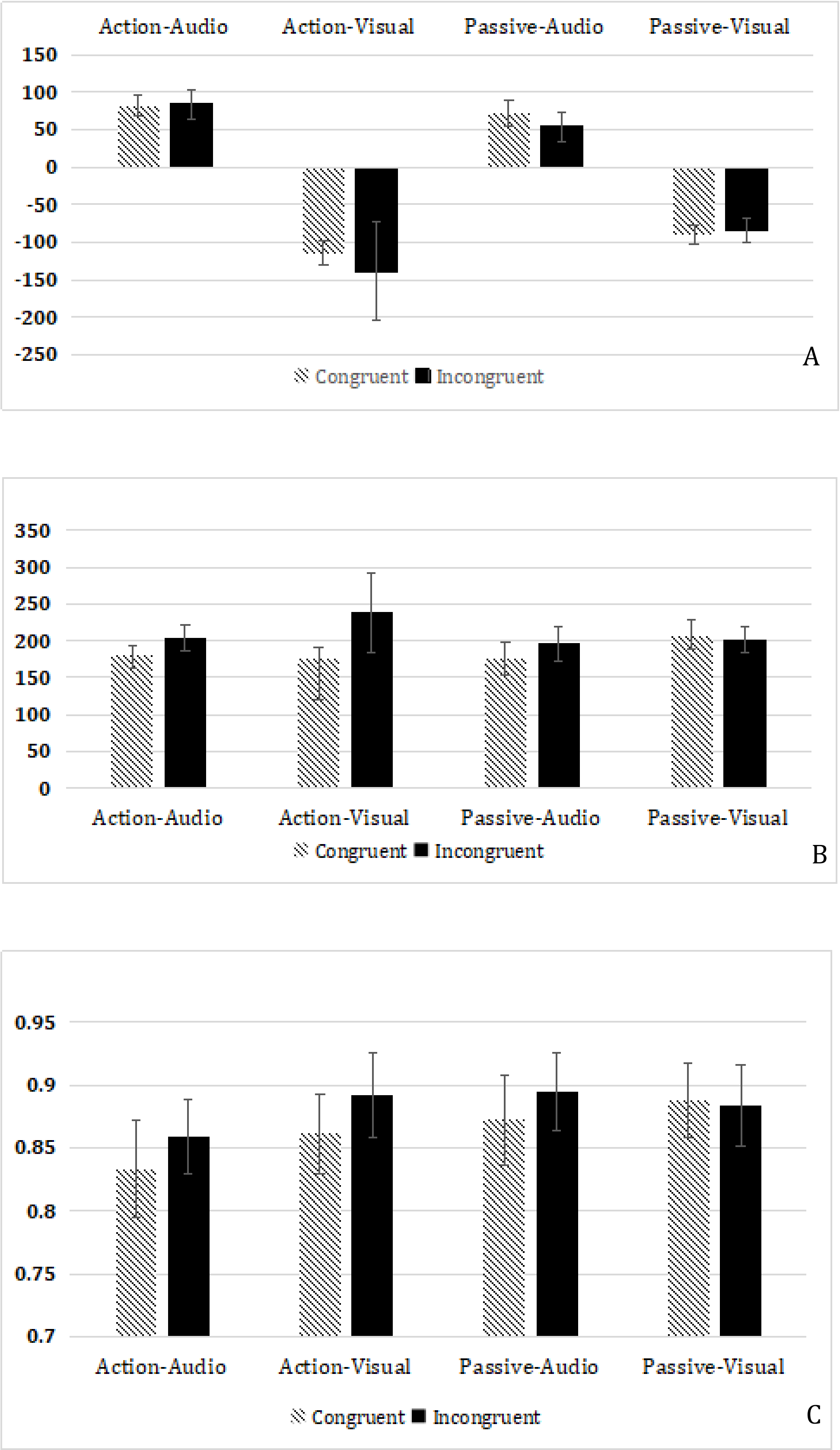


Figure 7. Mean values of A)PSS, B)SD and C)peak of probability for all experimental conditions tested.

3. General Discussion

In the present study, we intended to investigate how the action-effect delay and the semantic congruency of a multisensory action effect can influence the IB phenomenon. We attempted to address these two matters using for the first time multisensory naturalistic stimuli at a simultaneity judgment task, adjusting the methodology used by Thanopoulos et al., (2018). On both experiments, we failed to obtain the IB effect, since voluntary actions did not result in significant temporal shifts of the effect towards the actions when compared with passive conditions. Certain limitations due to sample sizes should be taken into account, as our effect sizes were quite small. Even so, our experiments systematically showed a significant impact of the temporal predictability of the effect (Experiment 1), as well as of the effect's particular characteristics (Experiment 2) on the perceived time of the effect occurrence by the participants.

Specifically, on Experiment 1 we hypothesized that voluntary actions followed by their effects at fixed short intervals (i.e., 250ms and 800ms) will lead to IB, while larger delays (i.e., 1000ms and 1250ms) will disrupt the phenomenon. Previous studies have reported contradicting findings regarding the induction of the phenomenon at various action-effect intervals, ranging from 100ms and up to 4s (Humphreys & Buehner, 2009; Engbert & Wohlschlager, 2007; Ruess et al., 2017), with most of them describing an increase on the temporal compression at 200-300ms and a constant decrease as the effect follows at greater intervals (Haggard et al, 2002; Haering & Kiesel, 2014; Ruess et al., 2017). We did not obtain significant difference on active trials as regards to the perceived time of the effect for any interval, thus we did not manage to replicate the Thanopoulos et al. (2108) study (who only utilized a 250ms action-effect interval). However, when the tone of the audiovisual pair appeared at fixed delays of 250ms, it was perceived significantly earlier than when it appeared at random intervals, revealing strong temporal binding. At this particular interval, participants also had a harder time to tell the stimuli apart when they appeared at fixed delays. These two findings likely underline the determining role of causality in temporal judgments, which is considered prominent by certain approaches to IB (Buehner, 2012; Buehner & Humphreys, 2009; Eagleman & Holcombe, 2002). According to this framework, actions are perceived as closer to their effects due to a general cause-effect binding process, on which voluntary action does not hold any special place when compared to other causes of events

(Buehner, 2012). Indeed, studies have reported temporal shifts at contexts where no voluntary action was present but the prerequisite of causality was satisfied, e.g., when participants observe other people perform an action that produced a certain effect (Engbert & Wohlschlager, 2007; Moore, Teufel, Subramaniam, Davis & Fletcher, 2013; Wohlschlager et al., 2003), on cases with shared actions (Strother, House & Obhi, 2010), when mechanical agents cause effects (Buehner, 2012), when actions are involuntarily produced (Arikan, Kemenade, Straube, Harris & Kircher, 2017), or when people are made to believe that they caused an action they did not (Desantis et al., 2011). Our data from Experiment 1 could likely be in accordance with such an interpretation, since both active and passive fixed conditions in our experiments maintain the same, complete causal sequence from the fixation image to the pair of image and sound that follows, regardless of the way the pair appears. This is further supported by the fact that the shift for 250ms happens only on the predictable interval, since causality between events is more likely inferred at short intervals (Ruess et al., 2017). However, it has also been shown that immediate causality is not a universal case, since humans tend to take into account previously learned delays between events and attribute causal relations according to them (Haering & Kiesel, 2015; Ruess et al., 2017), so we could speculate from the temporal shifts we observed that a certain causal link remains for intervals up to 1000ms but disappears at greater intervals, where the delay probably indicates two separate events that do not connect in some causal way. Furthermore, other studies have also reported that, since the causal relation is maintained, the shift can be preserved and even enlarged for intervals around 1s (e.g., Buehner & Humphreys, 2009).

At this point, we should also stress that the vast majority of studies which have reported a decrease of the temporal shifts as the action-effect delay increases made use of abstract stimuli (e.g., Haggard et al., 2002; Ruess et al., 2017). Taking into account the fact that the brain infers the causal relation between events by weighing the reliability of different cues in an optimal manner (Moore & Haggard, 2008), we could probably assume that, when abstract stimuli are utilized, parameters such as contiguity and temporal predictability are considered more reliable for a causal link formation. On the other hand, and this could be the case on our experiments, when stimuli have a specific semantic content, they can remain causally connected even if temporal parameters become distorted within certain limits. Specifically, for 800ms and 1000ms, we observed that the tone was perceived earlier in time regardless of its temporal predictability, a case different from the one we observed at the 250ms delay. This might have occurred because, on the short interval, low-level parceptual factors (e.g. action-effect perceptual grouping; Kawabe, Roseboom & Nishida, 2013) as well as temporal predictability might have been considered as more determinant for the causal relation between the action and the effect. On the other hand, when the audiovisual pair was further apart from fixation, causal relations could rather be determined by semantics, judged by the brain as more reliable. However, the temporal limits, within which a causal relation can be maintained based mainly on cognitive factors, despite of small low-level discrepancies, has to be further investigated. Our results, showing probable saturation of the effect at 1250ms, might point towards such a limit. Also, the specific contribution of low- and high-level factors in the determination of causality as action-effect intervals vary should be elucidated by future research.

Further support for this interpretation comes from our second experiment, in which we presented visual or auditory fixed effects, paired with naturalistic or abstract stimuli on the corresponding modality. Experiment 2 showed a clear effect of the modality in which the fixed effect was presented, with visual effects being perceived earlier than their actual occurrence, regardless of the presence of a voluntary action, while auditory effects were perceived later, especially when they were self-produced. This is indicative of the role of causality, which needed to be satisfied for the shifts to occur. Specifically, we used naturalistic visual stimuli that depicted the familiar sequence of a hand hitting a surface, in order to take advantage of the inherent causal relationship between those stimuli and enhance the causal relationship between the self-generated action and its outcome, adding validity from participants' everyday experience. On this particular sequence, given that the fixation was an image that showed the starting point of the sequence (a hand over a surface), it is highly probable that the complex visual outcome of the action (the picture which depicted the hand hitting the surface, i.e., the end of the sequence) was much more informative and indicative of the sequence's end than the complex auditory outcome on auditory blocks (the sound of a hand hitting a surface). Thus, the appearance of this particular visual outcome was necessary for the sequence to be completed and the blocks on which it always appeared (i.e., the visual-effect blocks) were the only ones that satisfied the predictability factor that has been previously shown to be important for determining causality and, thus, lead to IB induction (Cravo et al., 2011; Thanopoulos et al., 2018). On the other hand, although this particular visual stimulus also appeared on auditory blocks, it only appeared half of the times along with the fixed auditory effect, resulting in minimized predictability and partly a discontinuity to the causal sequence in general, since the auditory effect was apparently less informative as action outcome when appearing on its own. This was not the case on Experiment 1, when both naturalistic stimuli appeared together in all trials. Similarly, Thanopoulos et al. (2018) report minimum temporal shifts when the sequence with the same audiovisual effect (hitting hand image and sound) starts with a mismatching fixation image (foot over surface). Finally, the fact that the auditory fixed effect was perceived so far from its actual occurrence, especially when it was not self-generated, can possibly be explained by its matching (on half of the trials) with an abstract visual stimulus that was likely perceived as shorter in duration, thus perceived as ending first (see below).

Another interesting part of our findings regards the possible interactions between the modalities of the multisensory effect and the action, which could also partly explain why active and passive conditions did not differ at none of our experiments, not necessarily being mutually exclusive with the aforementioned approach. Given the strong semantic congruency of the audiovisual pair (the image and the impact sound of a hand hitting a surface) in all trials of Experiment 1 and on congruent conditions in Experiment 2, we had hypothesized that the two stimuli were likely unified in a single percept at some extent (i.e., unity assumption; Vatakis & Spence, 2007; Welch & Warren, 1980). Such perceptual integration of consistent stimuli has previously been reported to produce increased tolerance to temporal asynchronies (e.g., Arikan et al., 2017), or even temporal mislocations of the stimuli to satisfy the unified percept (Chen & Spence, 2017; Vatakis, 2013). In this case, it is quite possible that multisensory integration of the stimuli counteracted the effect of voluntary action to temporal perception, since the unification of the stimuli could be stronger than the attraction of the action to the fixed effect alone, resulting to a shift of the multisensory event as a whole. This would explain the significantly larger SDs on fixed trials in Experiment 1, since the temporally predictable tone was likely shifted along with the image. Parsons et al. (2013) on their attempt to disambiguate between the motor-recalibration and the IB account, predict that, when actions produce multisensory effects, humans will either recalibrate between the mismatching timings of different modalities, thus stimuli presented on different modalities will be independently shifted, or they will shift effects towards the action as a whole, due to a specific impact of intention on time perception. Although this research reaches a conclusion that favors the motor-recalibration hypothesis, our data could probably match to the second prediction. Yet, if this were the case, we would expect congruent outcomes in Experiment 2 to be more difficult to tell apart in the simultaneity judgment, a hypothesis not supported by our data, which showed significantly higher SD values for incongruent outcomes. This means participants found naturalistic stimuli more difficult to separate when they were paired with abstract ones and formed a semantically mismatched effect. Although surprising and seemingly contradicting to our initial hypothesis, this could be explained based on the informational content and complexity of the stimuli, which were not balanced. We cannot rule out that abstract, less informative stimuli were generally perceived as shorter in duration (e.g., Hogan, 1975, Zakay, 1993, where participants judged more complex stimuli as lasting longer), likely affecting the simultaneity judgment. An experimental setup, on which semantically congruent and incongruent stimuli will be equally complex (i.e., two complex images, semantically matching and mismatching a certain sound) would assist in drawing safer conclusions.

Concluding, our results offer some support to the hypothesis that causality is the determinant factor that binds actions to their effects, while also shedding some light in the interplay between multisensory integration and IB factors when it comes to time perception. Utilising familiar naturalistic stimuli, which hold an inherent causal link (Thanopoulos et al., 2018), allowed for certain manipulations of these factors in a more ecologically valid setup, closer to what humans experience when interacting with their everyday life environments. However, many aspects of the phenomenon are yet to be decisively addressed, while new questions arose from our results. The fact that the strong causal relation between our stimuli played a more determining role to our results than the presence or absence of a voluntary action contests the widely accepted notion in IB literature that intention is the decisive factor for the phenomenon. However, in order for this to be better investigated, our methodology should be modified to allow for a more subtle manipulation of factors that determine causality between events (e.g., semantics, temporal parameters and presence of volition), in order to decipher the particular contribution of each factor. Also, such an investigation should also be extended to examine how certain factors are differentially taken into account by the brain to infer causality as the interval between action and effect varies. Further, our methodology did not allow us to specify whether the multisensory effect of a voluntary action was shifted towards the action as a whole and what effect participants subjectively felt that they were causing with their action (i.e., the fixed effect alone or the whole audiovisual pair). A differentiated design could probably combine the implicit SJ task with an explicit agency judgment. to make this experience clearer.

References

Arikan, B., Kemenade, B., Straube, B., Harris, L., & Kircher, T. (2017). Voluntary and Involuntary Movements Widen the Window of Subjective Simultaneity. *i-Perception*, 8(4), 204166951771929.

Barlas, Z., & Kopp, S. (2018). Action Choice and Outcome Congruency Judgments. *Frontiers in Human Neuroscience*,12, 137.

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

Buehner, M. (2012). Understanding the Past, Predicting the Future. *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.

Chen, Y.-C. C., & Spence, C. (2017). Assessing the Role of the "Unity Assumption" on Multisensory Integration: A Review. *Frontiers in psychology*, 8, 445.

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Attention, perception & psychophysics*, 75(5), 790–811.

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

Desantis, A., Hughes, G., Waszak, F. (2012). Intentional binding is driven by the mere presence of an action and not by motor prediction. *PLoS One*, 7, e29557.

Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition: An International Journal*, 20(4), 1211-1220.

Desantis, A., Roussel, C., & Waszak, F. (2014). The temporal dynamics of the perceptual consequences of action-effect prediction. *Cognition*, 132, 243–250.

Desantis, A., Waszak, F., Moutsopoulou, K., & Haggard, P. (2016). How action structures time: About the perceived temporal order of action and predicted outcomes. *Cognition*,146, 100–9.

Ebert, J., & Wegner, D. (2010). Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition*, 19(1), 481–489.

Eagleman, D. M., & Holcombe, A. O. (2002). Causality and the perception of time. *Trends in Cognitive Sciences*, 6(8), 323-325.

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

Engbert, K., Wohlschlager, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1261−1268.

Engbert, K., Wohlschlager, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition*, 107(2), 693−704.

Guski, R., & Troje, N. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics*, 65(5), 789–800

Haggard, P. (2017). Sense of agency in the human brain. *Nature reviews Neuroscience*, 18(4), 196-207.

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), 382−385.

Haggard, P., Poonian, S., & Walsh, E. (2009). Representing the consequences of intentionally inhibited actions. *Brain Research*, 1286, 106–113.

Haering, C., & Kiesel, A. (2014). Intentional Binding is independent of the validity of the action effect's identity. *Acta Psychologica*, 152, 109–119.

Haering, C., & Kiesel, A. (2015). Was it me when it happened too early? Experience of delayed effects shapes sense of agency. *Cognition*, 136, 38–42

Hogan, H. (1975). Time perception and stimulus preference as a function of stimulus complexity. *Journal of Personality and Social Psychology*, 31(1), 32-35.

Humphreys, G., & Buehner, M. (2009). Magnitude estimation reveals temporal binding at super-second intervals. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1542-1549.

Kemenade, B., Arikan, B., Kircher, T., & Straube, B. (2016). Predicting the sensory consequences of one's own action: First evidence for multisensory facilitation. *Attention, Perception, & Psychophysics*, 78(8), 2515–2526.

Jensen, M., Vagnoni, E., Overgaard, M., & Haggard, P. (2014). Experience of action depends on intention, not body movement: An experiment on memory for mens rea. *Neuropsychologia*, 55, 122–127.

Jo, H.-G., Wittmann, M., Hinterberger, T., & Schmidt, S. (2014). The readiness potential reflects intentional binding. *Frontiers in Human Neuroscience*, 8, 421.

Kawabe, T., Roseboom, W., & Nishida, S. (2013). The sense of agency is action–effect causality perception based on cross-modal grouping. *Proc. R. Soc. B*, 280(1763), 20130991.

Libet, B., Gleason, C. A., Wright, Jr, 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, 623−642.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324.

Moore, J., Dickinson, A., & Fletcher, P. (2011). Sense of agency, associative learning, and schizotypy. *Consciousness and Cognition*, 20(3), 792–800.

Moore, J., & Fletcher, P. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, 21(1), 59–68.

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

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

Moore, J., Teufel, C., Subramaniam, N., Davis, G., & Fletcher, P. (2013). Attribution of Intentional Causation Influences the Perception of Observed Movements: Behavioral Evidence and Neural Correlates. *Frontiers in Psychology*, 4, 23.

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

Myung, I. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90–100.

Nolden, S., Haering, C., & Kiesel, A. (2012). Assessing intentional binding with the method of constant stimuli. *Consciousness and cognition*, 21(3), 1176–85.

Parsons, B., Novich, S., & Eagleman, D. (2013). Motor-Sensory Recalibration Modulates Perceived Simultaneity of Cross-Modal Events at Different Distances. *Frontiers in Psychology*, 4, 46.

Radeau, M., & Bertelson, P. (1974). The aftereffects of ventriloquism. *Quarterly Journal of Experimental Psychology*, 26, 63–71.

Rohde, M., Greiner, L., & Ernst, M. (2014). Asymmetries in visuomotor recalibration of time perception: Does causal binding distort the window of integration? *Acta Psychologica*,147, 127–135.

Rohde, M., Ernst, M. (2013). To lead and to lag-forward and backward recalibration of perceived visuo-motor simultaneity. *Frontiers in Psychology*, 3, 1-8.

Ruess, M., Thomaschke, R., & Kiesel, A. (2017). The time course of intentional binding. *Attention, Perception, & Psychophysics*, 79(4), 1123–1131.

Shams, L., & Beierholm, U. (2011). Humans' Multisensory Perception, from Integration to Segregation, Follows Bayesian Inference. *Sensory Cue Integration*, 251.

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, 117(11), 2341–2356.

Stetson, C., Cui, X., Montague, P., & Eagleman, D. (2006). Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation. *Neuron*, 51(5), 651–659.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition*, 17, 219–239.

Thanopoulos, V., Psarou, E., & Vatakis, A. (2018). Robust intentional binding for causally-linked sequences of naturalistic events but not for abstract event sequences. *Acta Psychologica*, 190, 159–173.

Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, 149(4), 439–446.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. *The American psychologist*, 54(7), 480–92.

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

Vastano, R., Pozzo, T., & Brass, M. (2017). The action congruency effect on the feelings of agency. *Consciousness and Cognition,* 51, 212–222.

Vatakis, A. (2013). The Role of Stimulus Properties and Cognitive Processes in the Quality of the Multisensory Perception of Synchrony. In Handbook of Experimental Phenomenology, L. Albertazzi (Ed.).

Vatakis, A., Bayliss, L., Zampini, M., & Spence, C. (2007). The influence of synchronous audiovisual distractors on audiovisual temporal order judgments. *Perception & psychophysics*, 69, 298–309.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Experimental Brain Research*, 185, 521–529.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & psychophysics*, 69(5), 744–56

Zakay, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics*, 54(5), 656–664.

Zampini, M., Guest, S., Shore, D., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, 67, 531–544.