

Predicting the sensory consequences of one's own action: First evidence for multisensory facilitation

Bianca M. van Kemenade¹ · B. Ezgi Arikan¹ · Tilo Kircher¹ · Benjamin Straube¹

© The Psychonomic Society, Inc. 2016

Abstract Predicting the sensory consequences of our own actions contributes to efficient sensory processing and might help distinguish the consequences of self- versus externally generated actions. Previous research using unimodal stimuli has provided evidence for the existence of a forward model, which explains how such sensory predictions are generated and used to guide behavior. However, whether and how we predict *multisensory* action outcomes remains largely unknown. Here, we investigated this question in two behavioral experiments. In Experiment 1, we presented unimodal (visual or auditory) and bimodal (visual and auditory) sensory feedback with various delays after a self-initiated buttonpress. Participants had to report whether they detected a delay between their buttonpress and the stimulus in the predefined task modality. In Experiment 2, the sensory feedback and task were the same as in Experiment 1, but in half of the trials the action was externally generated. We observed enhanced delay detection for bimodal relative to unimodal trials, with better performance in general for actively generated actions. Furthermore, in the active condition, the bimodal advantage was largest when the stimulus in the task-irrelevant modality was not delayed—that is, when it was time-contiguous with the action—as compared to when both the task-relevant and task-irrelevant modalities were delayed. This specific enhancement for trials with a nondelayed task-irrelevant modality was absent in the passive condition. These results suggest that a forward model creates predictions for multiple modalities, and consequently

contributes to multisensory interactions in the context of action.

Keywords Multisensory processing · Perception · Action

Theories on sensorimotor control have suggested that when we perform a voluntary action, we predict the sensory consequences of that action (Miall & Wolpert, 1996; Wolpert & Flanagan, 2001; Wolpert, Ghahramani, & Jordan, 1995). For example, when we reach for our cup of coffee, we expect to see our hand move in the planned direction. Similarly, when we want to say something, we expect to hear the sound of our voice. The theory of a “forward model” has been proposed as the underlying mechanism of action predictions, which uses a copy of the motor command, the efference copy, to generate predictions about the sensory outcome of our action (Miall & Wolpert, 1996; von Holst & Mittelstaedt, 1950/1971; Wolpert et al., 1995). This prediction is then compared with the actual sensory feedback. When the sensory feedback matches our prediction, the sensory action effects often generate less cortical response than externally generated stimuli (e.g., Blakemore, Wolpert, & Frith, 1998; Matsuzawa et al., 2005; Shergill et al., 2013). It is thought that such sensory attenuation may help us to distinguish stimuli generated by our own actions from those generated by the outside world, which are not associated with sensory attenuation (Blakemore, Wolpert, & Frith, 2002). In the case of a mismatch—due to, for instance, temporal or spatial violations—a prediction error is generated (Wolpert & Flanagan, 2001). Such prediction errors can be used to update our predictions and guide our actions, thus forming an integral part of motor learning.

Previous research has focused on predictive mechanisms in one modality only. For example, visual predictions have often been studied using joysticks controlling a cursor, or natural

✉ Bianca M. van Kemenade
vankemen@med.uni-marburg.de

¹ Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann-Straße 8, 35039 Marburg, Germany

feedback of the hand using a camera, with either temporal or spatial deviations of the feedback (Farrer, Bouchereau, Jeannerod, & Franck, 2008; Hoover & Harris, 2012; Leube et al., 2003). Actively producing speech and listening to recorded speech stimuli have been used to assess auditory prediction errors (Ford, Gray, Faustman, Heinks, & Mathalon, 2005; Wang et al., 2014). Finally, tactile stimuli with various delays have been used to investigate predictive mechanisms in the tactile domain (e.g., Blakemore, Wolpert, & Frith, 1998). However, many of our actions have *multisensory* consequences. For example, knocking on a door will not only produce a sound, but will also involve seeing the hand move and feeling the impact of the fingers on the door, besides receiving proprioceptive feedback. As previous research has shown, stimuli from different modalities can influence and often facilitate the perception of each other (e.g., Ernst & Banks, 2002; McDonald, Teder-Sälejärvi, & Hillyard, 2000; for a review, see Alais, Newell, & Mamassian, 2010; Driver & Noesselt, 2008; Ernst & Bühlhoff, 2004). However, it is unclear whether multisensory predictions are made when performing an action. The forward model has proven to be a useful model in understanding perception and action, at least for explaining the prediction of unisensory action outcomes. Investigating the prediction of multisensory action outcomes will add further knowledge on perception and action in a multisensory and thus more natural context.

Although most studies suggest that a match between predicted and actual action outcomes leads to sensory attenuation, some studies suggest otherwise. For example, Shimada et al. (2010) found a higher sensitivity to detect delays between action and feedback for voluntary than for externally generated actions. Similarly, an EEG study uncovered increased visual P1 amplitudes for action-triggered visual stimuli, as compared to passively viewed stimuli (Hughes & Waszak, 2011). For auditory stimuli, Reznik, Henkin, Levy, and Mukamel (2015) found that participants not only were more sensitive to self-generated sounds on a behavioral level, but also showed enhanced brain activity in auditory cortex during self-generated relative to externally generated sounds. Similarly, Ackerley and colleagues (2012) found that active self-touch led to higher activity in sensorimotor areas. Voluntary action thus seems to be able to either suppress or enhance sensory processing of action outcomes. The results of several studies suggest that the effect of voluntary action depends on various contextual factors. For example, Reznik et al. (2015) found that sensory attenuation only occurs when the sounds that resulted from the action were loud; when the sounds were of much less intensity, enhancement of perceived loudness was found. Furthermore, Hoover and Harris (2012) found increased detection of delays between action and outcome when the view of the hand was natural. When the visual feedback of

the hand was flipped—that is, when it showed unnatural feedback of the hand being upside down—detection was worse.

In the present study, two experiments were performed to investigate temporal predictions of multisensory action consequences. Here we refer to temporal prediction as a form of action–stimulus contiguity and not as a learned prediction based on the frequency of presented delays throughout the experiments. In Experiment 1, voluntary actions elicited the presentation of stimuli with variable delays in the visual and auditory modality. In Experiment 2, the paradigm was repeated with both self-generated (active) and externally generated (passive) actions. In both experiments, participants had to detect delays between action and feedback. Since most actions lead to multisensory consequences, we hypothesized that our internal forward model predicts multisensory consequences of action, rather than only task-relevant unisensory consequences. As we mentioned before, Hoover and Harris showed that receiving more natural feedback of the action enhanced the detection of delays between action and outcome (Hoover & Harris, 2012). Since we hypothesized that it is more natural to predict multisensory action consequences, we expected that receiving bimodal audiovisual feedback would enhance the detection of delays as compared with unimodal feedback. Furthermore, we expected better performance when one modality is time-contiguous with the action, as compared to a condition with multiple prediction errors (i.e., when temporal predictions of both modalities are violated). We assumed that these results would be abolished in the passive condition, in which the actions were externally generated in an unpredictable manner.

It should be noted that various factors play roles in the difference between self-generated and externally generated stimulus processing. The first would be the presence and absence of efference copy signals, which may have attenuating or enhancing effects as detailed above. However, the efference copy is not the sole factor that is involved: There are also often differences in temporal control between self-generated and externally generated stimuli. When stimuli are self-generated, anticipation of the action allows for better temporal prediction than is associated with externally generated stimuli, especially when these are presented in an unpredicted manner (Hughes et al., 2013). Unpredictable events may attract participants' attention, whereas when we produce the stimuli ourselves, the enhanced readiness may enable us to reduce distraction and focus our attention more on the upcoming stimuli. Further evidence that the efference copy is not the only defining factor of voluntary action comes from studies on temporal control (Cravo, Claessens, & Baldo, 2011), causal beliefs (Desantis et al., 2016), and the influence of passively induced vestibular signals (Durgin, Gigone, & Scott, 2005). Here, it is not our aim to specify the exact source of differences between active and passive actions, but rather, to investigate

differences in processing unimodal versus bimodal action consequences generated either actively or passively.

Experiment 1

Method

Participants Twenty-four healthy, right-handed (Edinburgh Handedness Inventory) participants with normal or corrected-to-normal vision took part in the experiment. Based on their performance, one participant had to be excluded (see the Analysis section below), resulting in a final sample of 23 participants (11 male, 12 female; age 19–35, mean age = 24.5 years). The experiment was approved by the local ethics committee in accordance with the Declaration of Helsinki.

Stimuli and procedure The participants were tested in a quiet, dimly lit room, where they sat behind a computer screen (60 Hz) at a viewing distance of 54 cm. They placed their right hand on a button pad, with their right index finger touching the button. The button pad was located inside a black box, so that participants could not see their hand. During the experiment, the participants were wearing headphones, through which auditory stimuli were delivered in the form of a beep, composed of a 250 Hz sine wave. Furthermore, white noise was presented throughout the whole experiment, to mask the sound of the buttonpresses. The visual stimulus was a black dot (1.5° of visual angle), presented centrally on a medium gray background. Stimuli were presented using Octave and the Psychophysics Toolbox (Brainard, 1997). A chin rest was used to stabilize the participants' head during the experiment.

The participants had to perform buttonpresses with their right index finger, which would elicit the appearance of either a dot on the screen, or a tone, or both. The stimuli were presented either at the time of the buttonpress or with a variable delay. The participants' task was to detect delays between their buttonpress and the presented stimuli. They answered "Yes, there was a delay" by pressing a button with their left middle finger, or "No, there was no delay" by pressing a button with their left index finger for "No." Participants always had to report the delays in only one modality, referred to as the "task modality" in this article. Furthermore, in bimodal trials participants only had to report whether they detected a delay between their action and the task modality; that is, the other modality (referred to as the "task-irrelevant modality") was not important for the task. Participants were instructed at the start of the run which modality was the task modality in that particular run, so they knew which modality to focus on. Apart from bimodal trials, each run contained only unimodal trials of the respective task modality; that is, visual task runs only contained visual unimodal trials, auditory task runs only auditory unimodal trials. In unimodal trials, the delay between

action and stimulus was one of the six predefined delays (0, 83, 167, 250, 333, or 417 ms, presented in frames [0, 5, 10, 15, 20, or 25 frames]). In bimodal trials, the task modality was also presented with one of these six delays, the task-irrelevant modality with one of three delays (0, 167, or 417 ms). Unimodal and bimodal trials were randomized within each run. Unimodal stimuli were presented for 1 second. Bimodal stimuli were also presented for 1 s if both had the same delay after the buttonpress. However, in trials in which one stimulus was presented later than the other, the first stimulus was presented for 1 s, and the second only until the presentation duration of this first stimulus was reached. Both stimuli thus disappeared at the same time, regardless of their individual onset time. This was done to avoid that participants could use the offset of the stimuli to perform the task, as we specifically wanted them to judge the delay between action and the onset of the stimuli.

The procedure during a trial was as follows (see Fig. 1). Each trial started with an intertrial interval (1, 1.5, or 2 s) with a fixation cross, after which a cue appeared in the form of the outline of a square (3.2° of visual angle) surrounding the fixation cross. This cue indicated that from now on, participants could press the button with their right index finger. They were instructed to wait with their buttonpress for at least 700 ms after the appearance of the cue and to wait as long as they wanted. This was done to elicit a well-prepared, self-initiated buttonpress, rather than an automatic action as a reflex to the cue (Rohde & Ernst, 2013). If the button was pressed too early, the text "Too early" was presented and the trial was repeated. When the button was pressed at least 700 ms after the cue, the sensory feedback was presented. After offset of the stimuli, a 500-ms interval with a fixation cross followed. Subsequently, the question "Delay? Yes/No" was presented on the screen. Participants were instructed to be as accurate as possible, and were not required to be as fast as possible. They were given a maximum of 4 s to answer.

Prior to the experiment, participants were familiarized with the paradigm. First, they could press the button several times to see delayed (417 ms) and nondelayed feedback. Then, to become familiar with the paradigm, they completed a short training (20 trials) with the same procedure as the main experiment. Only the delays of 0 and 417 ms were included in this training, to enhance the difference between trials with delay and trials without delay, since the smaller delays would potentially be difficult to detect when they were encountered for the first time. During the training, feedback about participants' performance (correct or incorrect) was given. No feedback was given during the main experiment. The main experiment comprised 540 trials in total: We presented ten trials for each delay, thus leading to 60 unimodal visual trials and 60 unimodal auditory trials. Furthermore, for the bimodal condition with the visual task modality, we presented 60 trials with a nondelayed auditory modality, 60 trials with the auditory

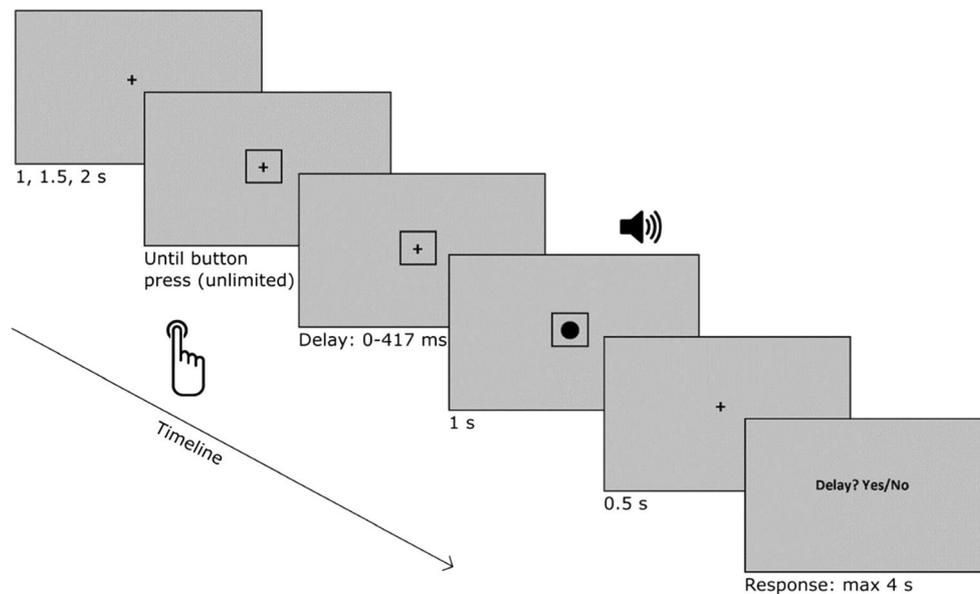


Fig. 1 Example of a bimodal trial. Participants had to wait with their buttonpress until the cue appeared, and they could take as much time as they wanted. After a variable delay, unimodal or bimodal stimuli were

presented. Participants had to report whether they detected a delay between their buttonpress and the stimulus of the task modality

stimulus delayed by 167 ms, and 60 trials with the auditory stimulus delayed by 417 ms. The same numbers of trials were used for the bimodal condition with the auditory task modality. In this way, the visual and auditory stimuli were congruent with each other (i.e., they had the same delay) for the delays of 0, 167, and 417 ms. To balance out the potential effects of stimulus congruency, we also added congruent bimodal trials for the delays of 83, 250, and 333 ms, with ten trials for each delay and each task modality. With these additional 60 trials, the total number of trials added up to 540. The experiment was divided into four runs: two visual-task-modality runs and two auditory-task-modality runs. Each run thus comprised 135 trials, consisting of both unimodal and bimodal trials with the respective task modality. White noise was played throughout the whole experiment to mask the sound of the buttonpress.

Analysis Using Psignifit (Fründ, Haenel, & Wichmann, 2011) in MATLAB, logistic psychometric functions were fitted to the data. The points of subjective equality (PSE) and the slopes of the functions were used to analyze the differences between the conditions with repeated measures analyses of variance (ANOVAs). The PSE reflects the detection threshold at which 50% of delays are detected. Thus, lower PSE values reflect better detection performance. Since proper PSE estimation requires a performance of at least 50%, participants who never reached a performance of at least 50%, even in the condition with the longest delay (417 ms), were excluded. One participant (from an original number of 24 participants) had to be excluded because of this criterion.

Repeated measures ANOVAs using SPSS were performed on the PSEs and slope values, which were extracted from each participant individually. In the first analysis, unimodal trials were compared to all bimodal trials together. The second analysis focused on the bimodal trials, in which the effect of delay in the second, task-irrelevant modality was investigated. When Mauchly's test indicated that the assumption of sphericity had been violated, a Greenhouse–Geisser correction was applied. Post-hoc *t* tests (Bonferroni-corrected) were conducted to verify the directions of the effects.

Results

Unimodal versus bimodal Figure 2 depicts the mean delay detection performance as a function of delay, averaged across all participants. A repeated measures ANOVA performed on the PSEs, using the factors Feedback Type (unimodal vs. bimodal) and Task Modality (visual vs. auditory), revealed significant main effects of feedback type and task modality [$F(1, 22) = 31.242, p < .001, \eta_p^2 = .587$, and $F(1, 22) = 4.628, p = .042, \eta_p^2 = .175$, respectively; see Fig. 3], with lower PSEs for bimodal than for unimodal trials, and lower PSEs for visual than for auditory trials. The interaction between these factors was not significant [$F(1, 22) = 3.294, p = .083, \eta_p^2 = .130$]. Analysis of the slopes did not show any significant main effects [condition, $F(1, 22) = 1.201, p = .285, \eta_p^2 = .052$; task modality, $F(1, 22) < 0.001, p = .993, \eta_p^2 < .001$]. The Feedback Type \times Task Modality interaction also did not reach significance [$F(1, 22) = 1.257, p = .274, \eta_p^2 = .054$].

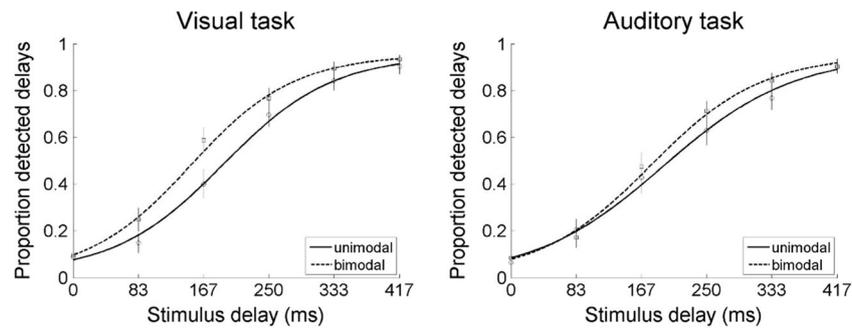


Fig. 2 Group psychometric functions for unimodal and bimodal trials, plotted over the averaged data points. Note that these curves are fitted over averaged data for illustration purposes; the actual analyses were done on individually fitted curves

Bimodal delays For each task modality, the task-irrelevant modality was presented with three different delays (0, 167, or 417 ms). These three delay conditions were investigated in a repeated measures ANOVA, using Delay (the three delay conditions of the task-irrelevant modality) and Task Modality (visual vs. auditory) as factors. Mauchly’s test indicated that the assumption of sphericity had been violated for delay [$\chi^2(2) = 7.737, p = .021$], so a Greenhouse–Geisser correction was applied. The analysis of the PSEs showed significant main effects of delay and task modality [$F(1.529, 33.6344) = 11.503, p < .001, \eta_p^2 = .343$, and $F(1, 22) = 11.355, p = .003, \eta_p^2 = .340$, respectively; see Fig. 4], but no significant interaction [$F(2, 44) = 2.069, p = .138, \eta_p^2 = .086$]. To explore the main effect of delay, post-hoc paired *t* tests were performed on the different delays, collapsed across task modalities. These showed that the PSEs in trials with a nondelayed task-irrelevant modality were significantly lower than the PSEs of trials with an intermediate [$t(22) = -4.920, p < .001, d = -0.580$] or long [$t(22) = -3.255, p = .004, d = -0.510$] delay in the task-irrelevant modality; see Fig. 5. The PSEs of trials with an intermediate delay were not significantly different from those with a long delay [$t(22) = 0.221, p = .827, d = 0.020$]. For the slopes, no significant main effects were found [delay, $F(1.203, 26.463) = 1.766, p = .183, \eta_p^2 = .074$; task modality,

$F(1, 22) = 0.001, p = .972, \eta_p^2 < .001$]. The Delay \times Task Modality was also not significant [$F(1.491, 32.796) = 0.652, p = .484, \eta_p^2 = .029$].

These results suggest that a stimulus order effect could have influenced the results, since in the condition with a nondelayed task-irrelevant modality, the task modality was always presented *after* the task-irrelevant modality. Therefore, we performed post-hoc comparisons of the unimodal conditions with the bimodal conditions with delays of 0 and 417 ms. These comparisons revealed that in both bimodal conditions the thresholds were significantly lower than in the unimodal conditions, for both visual [$t(22) = 5.785, p < .001, d = 1.104$, and $t(22) = 2.377, p = .027, d = 0.329$, respectively] and auditory [$t(22) = 4.623, p < .001, d = 0.574$, and $t(22) = 2.347, p = .028, d = 0.229$, respectively] trials. In other words, bimodal facilitation was found, regardless of the order of presentation of the task-relevant and task-irrelevant modalities.

Experiment 2

Method

Participants A total of 24 healthy, right-handed (Edinburgh Handedness Inventory) with normal or corrected-to-normal vision took part in the experiment. Due to technical problems with the passive button device, the data from one participant had to be excluded. Another participant had to be excluded due to low performance (see the Analysis section of Exp. 1), resulting in a final sample of 22 participants (11 male, 11 female; ages 19–30, mean age = 25.1 years). The experiment was approved by the local ethics committee in accordance with the Declaration of Helsinki. Two participants had participated before in Experiment 1; all the other participants were naïve to the study.

Apparatus and procedure The stimuli, procedure, and analyses were similar to those aspects of Experiment 1, except for the following details. The experiment was now divided into

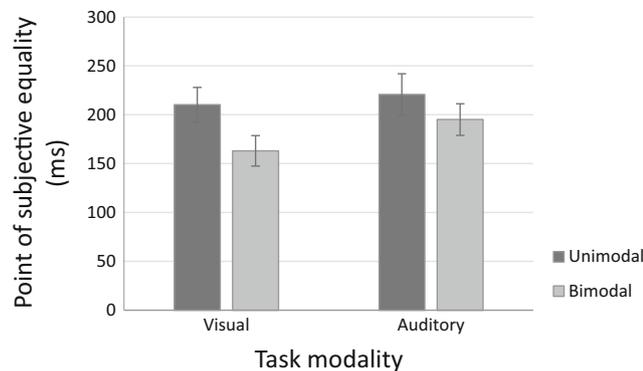


Fig. 3 Average points of subjective equality for unimodal and bimodal trials in both tasks. Significant main effects of feedback type (unimodal/bimodal) and task modality (visual/auditory) were found. Error bars represent standard errors of the means (SEMs)

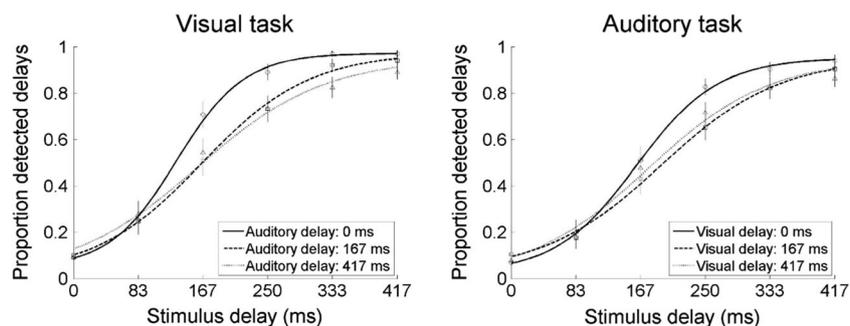


Fig. 4 Group psychometric functions for all bimodal conditions. Note that these curves are fitted over averaged data for illustration purposes; the actual analyses were done on individually fitted curves. In both tasks, the

point of subjective equality of trials in which the task-irrelevant modality was not delayed was significantly lower than when this modality was delayed

two active and two passive blocks, and the task modality was always the visual modality. A custom-made device with an electromagnet was used as a button pad throughout the experiment. In active conditions, the button was pressed actively by the participants. During the passive condition, the button was pulled down by the electromagnet, which was controlled by the computer. Each participant’s right index finger was loosely tied to the button with a soft bandage, so that in the passive conditions the finger would be pulled down with the button. In the active conditions, the bandage stayed tied, so that the same tactile information was present in all conditions. Participants wore earplugs, and white noise was played throughout the whole experiment to mask the sound of the electromagnet pulling the button down.

In both the active and passive blocks, participants had to wait for the appearance of a cue in the form of a square surrounding the fixation cross. In the active blocks, this cue indicated that participants could press the button from this moment onwards. In the passive blocks, this cue indicated that from now on, the button could be pulled down by the computer. The time between the cue and the passive buttonpress was jittered (0.5–3.5 s). Both active and passive buttonpresses elicited the presentation of a dot on the screen after a variable delay. In bimodal trials, a tone was added with a delay of either 0, 167, or 417 ms. Participants

were always required to report whether they had detected a delay between the buttonpress (active or passive) and the visual stimulus. As in Experiment 1, participants were familiarized with the stimuli prior to the actual experiment. First, they could press the button several times in order to see the delayed (417 ms) and nondelayed feedback. Furthermore, the button was pulled down by the computer a few times to show the delayed (417 ms) and nondelayed feedback after a passive buttonpress. Then, to become familiar with the paradigm, they completed a short training (20 trials) with the same procedure as the main experiment, during which feedback about their performance was presented. In the main experiment, no feedback was given. The main experiment consisted of 540 trials, just as in Experiment 1, with the same numbers of trials per condition. However, in Experiment 2 the auditory-task trials were replaced by passive trials.

Results

Unimodal versus bimodal Figure 6 depicts the mean delay detection performance as a function of delay, averaged across all participants. A repeated measures ANOVA performed on the PSEs, using the factors Feedback Type (unimodal/bimodal) and Action (active/passive), revealed significant main effects of feedback type and action [$F(1, 21) = 47.916, p < .001, \eta_p^2 = .695$, and $F(1, 21) = 14.453, p = .001, \eta_p^2 = .408$, respectively; see Fig. 7]. The interaction between these factors was not significant [$F(1, 21) = 0.007, p = .932, \eta_p^2 < .001$]. Analysis of the slopes showed a marginally significant effect of action [$F(1, 21) = 4.039, p = .057, \eta_p^2 = .161$]. We found no significant main effect of condition [$F(1, 21) = 0.032, p = .861, \eta_p^2 = .001$] or interaction [$F(1, 21) = 0.299, p = .59, \eta_p^2 = .014$].

Bimodal delays In both the active and passive conditions, the auditory, task-irrelevant modality was presented with three different delays (0, 167, or 417 ms). These three delay conditions were investigated in a repeated measures ANOVA, using Delay (the three delay conditions of the auditory modality)

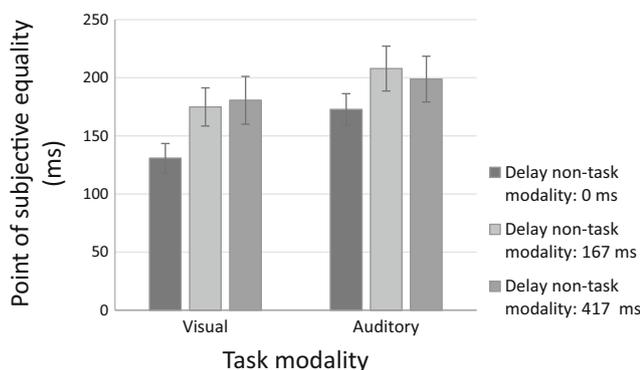


Fig. 5 Average points of subjective equality for bimodal trials per task. Significant main effects of delay and task modality were found

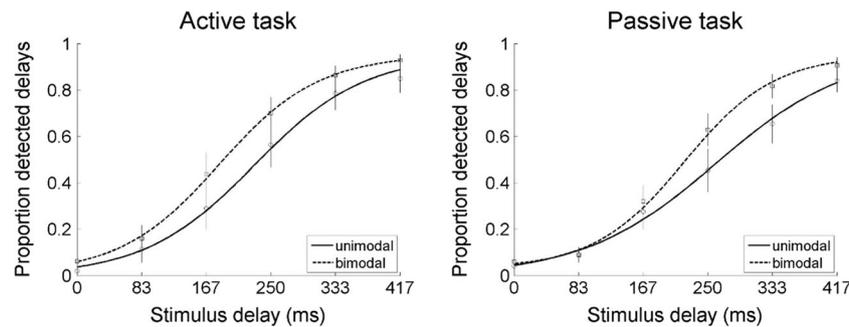


Fig. 6 Group psychometric functions for unimodal and bimodal trials, plotted over the averaged data points. Note that these curves are fitted over averaged data for illustration purposes; the actual analyses were

done on individually fitted curves. In both tasks, bimodal trials showed a significantly lower point of subjective equality than unimodal trials

and Action (active vs. passive) as factors. Figure 8 depicts the mean delay detection performance as a function of delay, averaged across all participants. The analysis of the PSEs showed significant main effects of both delay and action [$F(2, 42) = 10.218, p < .001, \eta_p^2 = .327$, and $F(1, 21) = 25.965, p < .001, \eta_p^2 = .553$, respectively]. Furthermore, the analysis revealed a significant Delay \times Action interaction [$F(2, 42) = 12.442, p < .001, \eta_p^2 = .372$]. For the slopes, a significant main effect of delay emerged [$F(1.128, 23.686) = 9.929, p = .003, \eta_p^2 = .321$], but no significant main effect of action [$F(1, 21) = 1.602, p = .219, \eta_p^2 = .071$] or Delay \times Action interaction [$F(1.167, 24.497) = 1.662, p = .212, \eta_p^2 = .073$]. Post-hoc paired t tests (Bonferroni-corrected for active and passive separately; $\alpha = .0167$; see Fig. 9) revealed a replication of the previous findings in the active condition: The PSE was significantly lower when the task-irrelevant modality was not delayed, as compared to when it was delayed by 167 ms [$t(21) = -3.914, p = .001, d = -0.678$] or 417 ms [$t(21) = -4.519, p < .001, d = -0.848$]. There was no difference between the delays of 167 and 417 ms [$t(21) = -0.725, p = .476, d = -0.088$]. In the passive condition, the PSEs were also significantly lower when the task-irrelevant modality was not delayed than when it was delayed by 167 ms [$t(21) = -2.629, p = .016, d = -0.334$]. However, the PSEs for a delay of

167 ms were significantly higher than those for the delay of 417 ms [$t(21) = 3.069, p = .006, d = 0.320$]. Importantly, we observed no difference when the task-irrelevant modality was not delayed versus when it was delayed by 417 ms [$t(21) = 0.125, p = .901, d = 0.017$]. The main effect of action showed that performance was generally worse in passive than in active trials. Post-hoc t tests on the differences between active and passive trials showed that the difference from the active condition was largest in the 0-ms condition; the farther the task-irrelevant stimulus was away from the action, the less enhancement was found [$\Delta_{\text{Act-pas}0\text{ms}} = -67.1 \text{ ms}, t(21) = -7.689, p < .001, d = -1.12$; $\Delta_{\text{Act-pas}167\text{ms}} = -38.2 \text{ ms}, t(21) = -3.675, p = .001, d = -0.40$; $\Delta_{\text{Act-pas}417\text{ms}} = -2.9 \text{ ms}, t(21) = -0.255, p = .801, d = -0.04$]. The specific advantage in delay detection for trials in which the task-irrelevant modality was not delayed versus any other delay thus seems to be specific to trials with voluntary action.

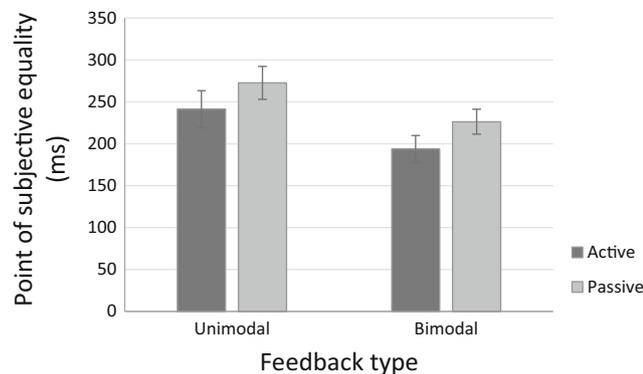


Fig. 7 Average points of subjective equality for unimodal and bimodal trials in both tasks. Significant main effects for feedback type and action were found. Error bars represent standard errors of the means (SEMs)

Buttonpress latencies In this experiment, we additionally recorded the times of the buttonpresses in each active condition, to investigate whether response latencies could have influenced the data. The average time that participants waited before pressing the button in all trials was 1,496 ms ($SD = 259$ ms; see Table 1 for the mean response latencies and standard deviations for each condition). Paired t tests revealed no significant differences in buttonpress latencies between the conditions, neither when comparing unimodal against all bimodal trials [$t(21) = 0.018, p = .986, d < 0.001$], nor when comparing the three bimodal conditions with each other [delay in the task-irrelevant modality: 0 vs. 167 ms, $t(21) = 0.735, p = .470, d = 0.078$; 0 vs. 417 ms, $t(21) = 0.400, p = .693, d = 0.031$; 167 vs. 417 ms, $t(21) = -0.659, p = .517, d = -0.051$].

Discussion

In the present study, we investigated the temporal prediction of multisensory consequences of one’s own action, using unimodal and bimodal visual and auditory stimuli, presented

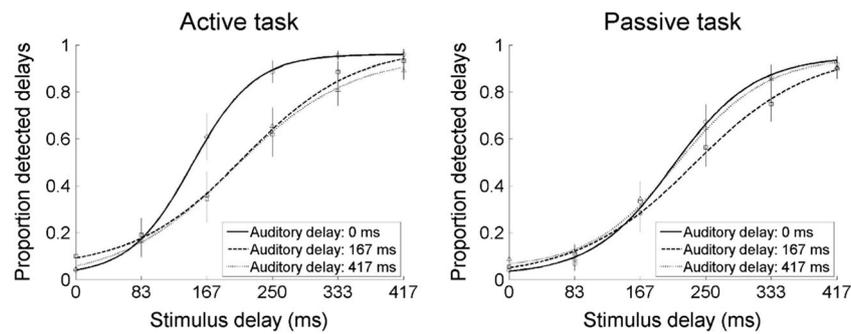


Fig. 8 Group psychometric functions for all bimodal conditions. Note that these curves are fitted over averaged data for illustration purposes; the actual analyses were done on individually fitted curves. In the active task,

at various delays after a buttonpress. In Experiment 1, in which the buttonpress was self-initiated, we observed an advantage for bimodal trials, as shown by a significant reduction in detection thresholds (PSEs). This advantage was especially evident when the task-irrelevant modality was time-congruent with the action. In Experiment 2, in which the buttonpress was either self-initiated (active) or externally generated (passive), we replicated these results for the active trials. In passive trials however, general performance was reduced as compared to the active condition. Importantly, we found a significant delay \times action interaction, with the largest difference in performance between active and passive trials when the task-irrelevant modality was time-contiguous with the action. Thus, we found a specific enhancement close to the action: the further the task-irrelevant stimulus was away from the action, the less enhancement it provided as compared to passive trials. These results suggest that the forward model generates predictions for multiple modalities: When a second, task-irrelevant, modality was present, performance was enhanced relative to when only the task modality was present. The enhancement was largest when the task-irrelevant modality was not delayed—that is, congruent with the action, and thus in line with the internal temporal prediction—as compared to when the temporal predictions for both modalities were violated. These results are specific to the active condition, which could

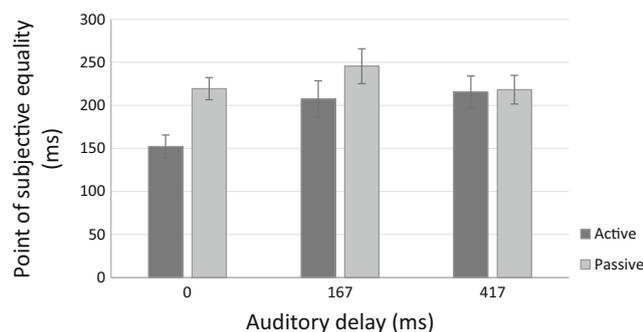


Fig. 9 Average points of subjective equality for bimodal trials per task. The difference between active and passive trials decreased with increasing delay of the task-irrelevant modality

the point of subjective equality of trials in which the task-irrelevant modality was not delayed was significantly lower than when this modality was delayed. This effect was abolished in the passive task

either mean that our results are due to efference copy mechanisms, or alternatively, that our effects are due to differences in temporal predictability of the stimuli. In active conditions, participants had a better temporal prediction about the upcoming stimuli, since they decided themselves when they pressed the button. In passive trials however, the button was pulled down at variable jittered time points, increasing the uncertainty about the temporal occurrence of the consequent stimuli. This difference in predictability reflects the natural characteristics of actively and passively generated stimuli in daily life, in which passively generated stimuli are generally less predictable than actively generated stimuli. Nevertheless, as has been described by Hughes et al. (2013), due to this predictability difference we cannot disentangle whether the effects are due solely to efference copy mechanisms, differences in temporal predictability, or both. In the present study, it was not our aim to specify the exact source of differences between active and passive actions, but rather, to investigate differences in processing unimodal versus bimodal action consequences generated either actively or passively. Which specific aspect of predictive mechanisms in voluntary action is responsible for our results would be an interesting topic for future studies.

Our results suggest that our internal model generates temporal predictions for at least two modalities. This is in line with previous studies with unisensory stimuli showing evidence for temporal predictions in the visual, auditory, and somatosensory system, however separately tested. For example, it has been widely accepted that presaccadic activity, a potential efference copy signal, predicts the visual consequences of saccades, thereby ensuring visual stability (e.g., Sommer & Wurtz, 2004; von Holst & Mittelstaedt, 1950/1971; but see Bridgeman, 2007). Furthermore, our internal model also plays an important role in predicting the visual consequences of more complex actions, such as various hand movements (Hoover & Harris, 2012; Knoblich & Kircher, 2004; Shimada, Qi, & Hiraki, 2010). Although many studies have focused on the role of the forward model in predicting visual consequences of actions, several studies have also shown its importance in predicting tactile (Blakemore et al.,

Table 1 Average buttonpress latencies for each condition of Experiment 2

	All Trials	Unimodal	Bimodal	Bimodal (Delay Task-Irrelevant Modality: 0 ms)	Bimodal (Delay Task-Irrelevant Modality: 167 ms)	Bimodal (Delay Task-Irrelevant Modality: 417 ms)
Mean (ms)	1,496	1,497	1,496	1,501	1,481	1,494
Standard deviation (ms)	259	263	262	272	273	249

No significant differences in buttonpress latencies between the conditions were found.

1998) and auditory (Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Ford et al., 2005) consequences. This suggests that predicting the outcome of our actions in different modalities is based on similar mechanisms. However, the mechanism behind the temporal prediction of *multisensory* consequences was unclear. To our knowledge, our study is the first to specifically investigate multisensory action predictions. Some studies have investigated the role of action expertise on the perception of multisensory actions, by showing drumming point-light displays with manipulated audio-visual information to experienced drummers and novices (Petrini et al., 2011; Petrini, Russell, & Pollick, 2009). They found that experienced drummers are better at detecting stimulus asynchronies than novices. Although these studies show that one can acquire internal models of action, which can aid in perceiving ambiguous multisensory information, they did not directly test the prediction of multisensory consequences of one's own action. One previous study showed that unpredicted visual stimuli affected loudness perception of auditory stimuli, both for self-generated stimuli and stimuli predicted by a cue (Desantis, Mamassian, Lisi, & Waszak, 2014). However, this study investigated the general cross-modal effect of predictability of task-irrelevant stimuli on the perception of the task stimuli. In our study, we were specifically interested in the temporal prediction of multisensory *action* consequences. Few other studies have included multisensory action consequences; however, they were used to study the sense of agency. For example, Farrer and colleagues found that the presentation of a sound at the time of the buttonpress significantly reduced the thresholds at which participants felt in full control of the appearance of the visual stimulus (Farrer, Valentin, & Hupé, 2013). Similarly, lower thresholds were found when additional tones were presented at the time of the buttonpress and visual stimulus in a cross-modal grouping paradigm with variable delayed visual stimuli (Kawabe, Roseboom, & Nishida, 2013). Although the sense of agency may rely, at least partly, on the same forward model used in sensorimotor predictions (Farrer, Frey, et al., 2008), it has also been suggested to emerge from an additional interpretive mechanism (Kawabe et al., 2013). In our study, we specifically investigated multisensory temporal predictions of the forward model by using a delay detection task instead of assessing the subjective experience of the action outcome.

Furthermore, our results go beyond these previous findings by manipulating the delay of *both* modalities, which revealed the additional effect that facilitation occurs specifically when one modality is congruent with the action. Nevertheless, a recent study by Kawabe (2015) had a quite similar paradigm. Although Kawabe's study was presented as investigating the sense of agency, the paradigm of their first experiment involved a delay detection task instead of an explicit agency judgment task. They recorded a hand movement and showed this visual feedback at various delays. Their results are partly in line with ours; in bimodal trials, delay detection was best when the task-irrelevant modality was not delayed. Nevertheless, no bimodal enhancement was found; performance in the bimodal condition with the nondelayed task-irrelevant modality was just as good as in the unimodal condition. Our and their bimodal effects are congruent, but our bimodal enhancement differs from their results. However, their study differs in some points from ours: first of all, Kawabe was interested in the general effect of delayed visual feedback on the sense of control over the auditory stimulus, ranging from no visual feedback to extremely delayed visual feedback, and thus had much less unimodal trials than bimodal trials. Furthermore, they showed recordings of the hand as visual feedback. Thus, their visual feedback was of the action itself, whereas the auditory stimulus was presented at or after the buttonpress. In our study, visual and auditory stimuli were both presented at or after the buttonpress. Importantly, our study included a passive condition in Experiment 2, in which the action was externally generated, whereas Kawabe's study did not include a passive condition. Thus, we partly replicate but also extend the findings from Kawabe's study by being able to link our results more closely to mechanisms involved specifically in voluntary actions.

Several studies on voluntary actions and their outcomes have observed a so-called "intentional binding effect" (Haggard, Clark, & Kalogeras, 2002). They found that participants judge the perceived time of their action closer to the time of the outcome, and the outcome closer to the action. In other words, the perceived times of action and outcome are attracted toward each other. For externally generated actions, the opposite effect can be seen. It has therefore been argued the forward model in voluntary actions helps to bind the outcome to our action to maintain a sense of agency (Haggard,

Clark, & Kalogeras, 2002). Although we did not specifically measure intentional binding, our results are in line with this phenomenon to a certain extent. First of all, we found bimodal enhancement, which has previously been found as well by Kawabe et al. (2013). Furthermore, for bimodal trials we see a specific advantage when the task-irrelevant modality is not delayed; with increasing delay of this stimulus, delay detection becomes worse. Similar effects have been found with unimodal action outcomes by Wen and colleagues, who found stronger intentional binding effects with increasing delays between action and outcome (Wen, Yamashita, & Asama, 2015). However, whereas the intentional binding effect is usually seen during voluntary actions, and repulsion of perceived time of action and outcome in externally generated actions, we found the opposite pattern: Delay detection performance was generally better in active than in passive conditions. In that sense, our results are more in line with various studies that have observed a similar enhancing effect in voluntary actions (Hoover & Harris, 2012; Shimada et al., 2010). The reason for this discrepancy cannot be resolved from our results. However, one should take into account that our paradigm differs in several aspects from traditional intentional binding studies. As we mentioned before, we cannot specify whether it is the efference copy or temporal control that can explain our results best—an issue present in intentional binding studies as well (Desantis, Hughes, & Waszak, 2012; Hughes, Desantis, & Waszak, 2013).

It should be taken into account that the order of our stimuli or the stimulus asynchrony could have played a role in our results. The bimodal advantage in the active conditions was largest when the stimulus in the task-irrelevant modality was not delayed—that is, congruent with the action. In these cases, the task modality comes after the task-irrelevant modality. Recognizing this asynchrony between the two stimuli may have helped the participants, as it indicates that the second stimulus, in this case the task stimulus, must be delayed. Indeed, humans can detect very short stimulus asynchronies (Spence, Shore, & Klein, 2001; Zampini, Shore, & Spence, 2003), suggesting that an order effect could play a role in this study. However, trials with a delayed task-irrelevant modality, in which case the task modality comes first, and trials with a nondelayed modality, in which the task modality comes after the task-irrelevant modality, both show significantly lower PSE values than unimodal trials. Importantly, in the passive condition, PSE values were lowest both when the task-irrelevant modality was not delayed and when it was delayed by 417 ms. In fact, the nondelayed condition and the condition with a delay of 417 ms did not differ at all. Thus, even if the order of the stimuli may have helped the participants, it cannot fully explain our results. First of all, multisensory facilitation was also found in the active task when the task modality was presented first. Secondly, the advantage for the nondelayed task-irrelevant modality was specific to the active task.

Another point that should be noted is that our stimuli had different durations in the bimodal conditions. We gave our stimuli the same offset, so the stimulus that was presented second had a shorter duration than the first stimulus, which always had a duration of 1,000 ms. Because we specifically wanted participants to judge the delays between buttonpress and the *onset* of the task stimulus, we decided to present the stimuli for this long duration and give the stimuli the same offset, so the only useful information to perform the task would be the onset of the stimuli, not the offset or their asynchrony. One could argue that different stimulus durations are not optimal. However, if the stimuli would have been presented for 1,000 ms each, the total duration of stimulus presentation would have varied, ranging from 1,000 to 1,417 ms, depending on the delay of the second stimulus. In addition, there would have been both an onset and an offset asynchrony, which participants could have used to enhance their performance. Therefore, we decided that it would be a better solution to keep the total stimulus duration the same, with the same offset for both stimuli. We cannot rule out completely that different stimulus durations would have had an effect on our data. However, Experiment 2 showed that different stimulus durations cannot fully explain our effects, since we do not see a specific enhancement for the 0-ms delay of the task-irrelevant modality in the passive condition.

Although the stimuli in the active conditions of our experiments should have been perceived as self-generated, one could argue that some stimuli with a long delay would be perceived as externally generated. This would mean that in some bimodal trials, one stimulus would be perceived as self-generated and the other as externally generated, which makes the results hard to interpret. However, it seems unlikely that some stimuli would have been considered externally generated. A study by Farrer et al. (2013) investigated the feeling of agency, and how this is influenced by various factors. Their main paradigm was similar to ours: Participants had to press a button, which would elicit the presentation of a dot on the screen after a variable delay. Instead of having to report whether they detected a delay, participants were asked to report whether they felt they were in control of the appearance of the dot, answering either with “full control,” “partial control” (“I detected a delay but still feel I was the agent”), and “no control.” Their results show that the feeling of “no control” arises after at least 628 ms or more, a much longer delay than our longest delay of 417 ms in all manipulations. Furthermore, participants in that study were told that they were not always the agent, and that the dot was sometimes controlled by the computer, making them perhaps even more prone to sometimes report “no control.” In our study, participants were explicitly told that they were always the agent. Taken together, it seems unlikely that some stimuli would be considered externally generated in our study.

Our results could also be explained by using the more general framework of predictive coding, or the free-energy principle. In this framework, the brain is regarded as a Bayesian inference machine, updating beliefs or expectations using sensory information (Friston, 2010; Picard & Friston, 2014). According to this theory, biological agents resist a natural tendency to disorder, requiring them to minimize the amount of surprise, or free energy, associated with encountered events. They manage to minimize free energy, or prediction error, through both perception and action. Perception can minimize prediction error by updating the predictions, action can do so through active inference, which changes actual sensory input to fit the predicted sensory input (Friston, 2010). An important distinction with the forward-model theory is that in the free energy formulation, movement is not driven by motor commands, but by predictions about the proprioceptive consequences of that movement. Thus, prediction errors are not there to refine motor behavior, but themselves *are* the motor signals (Friston, Daunizeau, Kilner, & Kiebel, 2010). In both theories, the efference copy plays an important role. However, whereas the forward model describes the efference copy as a copy from the motor signal, that is used to predict the sensory outcome of an action, the active inference account sees the efference copy as a bottom up prediction error signal used to optimize predictions. In our experimental approach, we specifically manipulated the efference copy presence by including active and passive conditions. Thus, comparing the forward model theory with the free energy principle is beyond the scope of this study, but would certainly be interesting for future studies.

Furthermore, future studies should transfer our findings to more natural contexts with a focus on realistic action outcomes (e.g., videotaped hand movements) and not only abstract action outcomes, as in our study. However, in a world in which we are surrounded by computers and other devices, it is a common action to press a button and expect a visual and/or auditory consequence, such as when typing a letter or playing a game. Thus, despite the setup being fairly abstract, it can still be considered ecologically valid.

All in all, we have shown that multisensory consequences of our own action lead to enhanced detection of delays between action and feedback. Furthermore, we have shown that timing matters: When the task-irrelevant modality is not delayed—that is, is time-contiguous with the action—performance is better than when the *temporal* predictions of both modalities are violated. This effect was specific to trials with voluntary action. Our findings point toward the idea that one forward model creates multisensory temporal predictions. Alternatively, it could be that separate forward models are used for each modality, which tightly interact. With our results, these two interpretations cannot be distinguished; future studies using functional magnetic resonance imaging or electroencephalography might be able to disentangle these two

possibilities. However, our study is an important first step in unraveling the nature of the temporal prediction of multisensory action consequences. The results support the idea that the forward model provides predictions for all modalities and consequently contributes to multisensory interactions in the context of action.

Author note All data are available at <http://dx.doi.org/10.5281/zenodo.59887>. This study was funded by the Deutsche Forschungsgemeinschaft (DFG) through the SFB/Transregio 135, “Cardinal Mechanisms of Perception: Prediction, Valuation, Categorization,” and through the International Research Training Group, IRTG 1901, “The Brain in Action–BrainAct.” B.S. is supported by DFG grants STR 1146/8-1 and STR 1146/9-1. We thank Jens Sommer and Kornelius Podranski for technical support, and Peer Herholz for help with data collection.

References

- Ackerley, R., Hassan, E., Curran, A., Wessberg, J., Olausson, J., & McGlone, F. (2012). An fMRI study on cortical responses during active self-touch and passive touch from others. *Frontiers in Behavioral Neuroscience*, *6*(51), 1–9. doi:10.3389/fnbeh.2012.00051
- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, *23*, 3–38. doi:10.1163/187847510X488603
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*, 635–640. doi:10.1038/2870
- Blakemore, S., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*, 237–242.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- Bridgeman, B. (2007). Efference copy and its limitations. *Computers in Biology and Medicine*, *37*, 924–929. doi:10.1016/j.compbimed.2006.07.001
- 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.
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*, 183–191.
- 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. doi:10.1371/journal.pone.0029557
- Desantis, A., Mamassian, P., Lisi, M., & Waszak, F. (2014). The prediction of visual stimuli influences auditory loudness discrimination. *Experimental Brain Research*, *232*, 3317–3324. doi:10.1007/s00221-014-4001-2
- 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*, 100–109.
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron*, *57*, 11–23. doi:10.1016/j.neuron.2007.12.013
- Durgin, F. H., Gigone, K., & Scott, R. (2005). Perception of visual speed while moving. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 339–353. doi:10.1037/0096-1523.31.2.339
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. doi:10.1038/415429a

- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169. doi:10.1016/j.tics.2004.02.002
- Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of agency. *Behavioural Neurology*, *19*, 53–57.
- Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S., & Grafton, S. T. (2008). The angular gyrus computes action awareness representations. *Cerebral Cortex*, *18*, 254–261. doi:10.1093/cercor/bhm050
- Farrer, C., Valentin, G., & Hupé, J. M. (2013). The time windows of the sense of agency. *Consciousness and Cognition*, *22*, 1431–1441. doi:10.1016/j.concog.2013.09.010
- Ford, J. M., Gray, M., Faustman, W. O., Heinks, T. H., & Mathalon, D. H. (2005). Reduced gamma-band coherence to distorted feedback during speech when what you say is not what you hear. *International Journal of Psychophysiology*, *57*, 143–150. doi:10.1016/j.ijpsycho.2005.03.002
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. doi:10.1038/nrn2787
- Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*, 227–260. doi:10.1007/s00422-010-0364-z
- Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11*(6), 16:1–19.
- Haggard, P., Clark, S., & Kalogerias, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*, 382–385. doi:10.1038/nn827
- Hoover, A. E. N., & Harris, L. R. (2012). Detecting delay in visual feedback of an action as a monitor of self recognition. *Experimental Brain Research*, *222*, 389–397. doi:10.1007/s00221-012-3224-3
- Hughes, G., & Waszak, F. (2011). NeuroImage ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *NeuroImage*, *56*, 1632–1640. doi:10.1016/j.neuroimage.2011.02.057
- Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*, 133–151. doi:10.1037/a0028566
- Kawabe, T. (2015). Delayed visual feedback of one's own action promotes sense of control for auditory events. *Frontiers in Integrative Neuroscience*, *9*(57), 1–7. doi:10.3389/fnint.2015.00057
- Kawabe, T., Roseboom, W., & Nishida, S. (2013). The sense of agency is action–effect causality perception based on cross-modal grouping. *Proceedings of the Royal Society B*, *280*, 20130991. doi:10.1098/rspb.2013.0991
- Knoblich, G., & Kircher, T. T. J. (2004). Deceiving oneself about being in control: Conscious detection of changes in visuomotor coupling. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 657–666. doi:10.1037/0096-1523.30.4.657
- Leube, D. T., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. T. J. (2003). The neural correlates of perceiving one's own movements. *NeuroImage*, *20*, 2084–2090. doi:10.1016/j.neuroimage.2003.07.033
- Matsuzawa, M., Matsuo, K., Sugio, T., Kato, C., & Nakai, T. (2005). Temporal relationship between action and visual outcome modulates brain activation: An fMRI study. *Magnetic Resonance in Medical Sciences*, *4*, 115–121.
- McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*, 906–908.
- Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*, 1265–1279. doi:10.1016/S0893-6080(96)00035-4
- Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, *110*, 432–439. doi:10.1016/j.cognition.2008.11.015
- Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., & Puce, A. (2011). Action expertise reduces brain activity for audiovisual matching actions: An fMRI study with expert drummers. *NeuroImage*, *56*, 1480–1492. doi:10.1016/j.neuroimage.2011.03.009
- Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, *83*, 1112–1118. doi:10.1212/WNL.0000000000000798
- Reznik, D., Henkin, Y., Levy, O., & Mukamel, R. (2015). Perceived loudness of self-generated sounds is differentially modified by expected sound intensity. *PLoS ONE*, *4*, e127651. doi:10.1371/journal.pone.0127651
- Rohde, M., & Ernst, M. O. (2013). To lead and to lag-forward and backward recalibration of perceived visuo-motor simultaneity. *Frontiers in Psychology*, *3*(599), 1–8. doi:10.3389/fpsyg.2012.00599
- Shergill, S. S., White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2013). Modulation of somatosensory processing by action. *NeuroImage*, *70*, 356–362.
- Shimada, S., Qi, Y., & Hiraki, K. (2010). Detection of visual feedback delay in active and passive self-body movements. *Experimental Brain Research*, *201*, 359–364. doi:10.1007/s00221-009-2028-6
- Sommer, M. A., & Wurtz, R. H. (2004). What the brain stem tells the frontal cortex: I Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *Journal of Neurophysiology*, *91*, 1381–1402. doi:10.1152/jn.00738.2003
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, *130*, 799–832. doi:10.1037/0096-3445.130.4.799
- von Holst, E., & Mittelstaedt, H. (1971). The principle of reafference: Interactions between the central nervous system and the peripheral organs. In P. C. Dodwell (Ed.), *Perceptual processing: Stimulus equivalence and pattern recognition* (pp. 41–72). New York, NY: Appleton-Century-Crofts (Original work published 1950).
- Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., & Ford, J. M. (2014). Action planning and predictive coding when speaking. *NeuroImage*, *91*, 91–98. doi:10.1016/j.neuroimage.2014.01.003
- Wen, W., Yamashita, A., & Asama, H. (2015). The influence of action–outcome delay and arousal on sense of agency and the intentional binding effect. *Consciousness and Cognition*, *36*, 87–95. doi:10.1016/j.concog.2015.06.004
- Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*, R729–R732. doi:10.1016/S0960-9822(01)00432-8
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882. doi:10.1126/science.7569931
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*, 198–210. doi:10.1007/s00221-003-1536-z