The virtual co-actor: the social Simon effect does not rely on online feedback from the other

Elize Vlainic[1], Roman Liepelt[2,3], Lorenza S. Colzato[1,4], Wolfgang Prinz[2] and Bernhard Hommel[1,4]*

[1] Cognitive Psychology Unit, Leiden University, Leiden, Netherlands
[2] Department of Psychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[3] Junior Group "Neurocognition of joint action", Department of Psychology, Westfälische Wilhelms-University, Münster, Germany
[4] Leiden Institute for Brain and Cognition, Leiden, Netherlands

The social Simon effect (SSE) occurs if two participants share a Simon task by making a Go/No-Go response to one of two stimulus features. If the two participants perform this version of the Simon task together, a Simon effect occurs (i.e., performance is better with spatial stimulus–response correspondence), but no effect is observed if participants perform the task separately. The SSE has been attributed to the automatic co-representation of the co-actor's actions, which suggests that it relies on online information about the other's actions. To test this implication, we investigated whether the SSE varies with the presence and amount of online action-related feedback from the other person. Experiment 1 replicated the SSE with auditory stimuli. Experiment 2, in which participants were blindfolded, demonstrated that visual feedback from the other's actions is not necessary for the SSE to occur. Experiment 3 replicated Experiment 2 with a regular and a soundless keyboard. A comparable SSE was obtained in both conditions, suggesting that even auditory online input from the other's actions is not necessary. Taken together, our data suggest that the SSE does not rely on online information about the co-actor's actions but that *a priori* offline information about another actor's presence is sufficient to generate the effect.

Keywords: Simon effect, social Simon effect, action representation, task representation, grounded cognition

INTRODUCTION

Humans are active agents who organize their behavior according to their plans and action goals. However, where those plans and goals come from and how they are acquired is not very understood. According to the ideomotor approach to voluntary action (Lotze, 1852; James, 1890; for an overview, see Stock and Stock, 2004), actions are cognitively represented in terms of their sensory consequences, so that the acquisition of action plans amounts to the experience-driven integration of motor patterns with codes of their sensory effects (Elsner and Hommel, 2001). Indeed, numerous studies have provided evidence that performing a movement creates associations between the underlying motor pattern and the sensory consequences that go along with executing this pattern (for an overview, see Hommel, 2009). This implies that our cognitive action representations are grounded in sensory experience, that is, in the perceptual consequences a given action was experienced to create. According to ideomotor theory, this perceptual grounding provides us with the means to carry out movements intentionally: we internally re-create the sensory experience of the action effects to some degree (in other words, we anticipate them) and thereby reactivate the associated motor pattern that will then produce the anticipated effects in the external world (Elsner and Hommel, 2001; Hommel, 2009).

Recent research has raised the possibility that action representations do not only comprise of information about the sensory consequences of one's own action but that information about other people's actions might also be considered. Most of this research

stems from the so-called "Social Simon paradigm," in which two persons share a Simon task (Sebanz et al., 2003). In the standard Simon task, single participants carry out left and right responses to a non-spatial attribute of stimuli that appear randomly on the left or right side. The standard finding in this task is that participants perform better if the stimulus happens to appear on the side of the correct response than if it does not (Simon and Rudell, 1967). Sebanz et al. (2003) had two participants share this task, so that each participant responded to only one of the stimuli by pressing a single key, which from the perspective of each participant rendered the task a Go/No-Go task. While performing this Go/No-Go version alone did not elicit a Simon effect, working on the task together with a co-actor did. This shared-task effect has been called the social Simon effect (SSE; Sebanz et al., 2003).

The SSE suggests that action or task representations are grounded not only in the experience of one's own actions but that they can also include aspects of the current social or at least situational context (Hommel et al., 2009), which seems to imply that action planning is truly situated (Clancey, 1997). Given that we are social animals used to act in social context, which often requires the consideration of other people's activities, this may not come as a surprise. However, the cognitive mechanisms responsible for integrating information about the current action context are not very well understood. According to Sebanz et al. (2003), the SSE might suggest that people do not only create cognitive representations of their own actions but they may also automatically co-represent the actions of a co-actor. In particular, Sebanz et al. (2003) suggest that

the effect may arise at a representational level that does not distinguish between one's own and another person's actions. According to the ideomotor principle (Hommel, 2009), both types of actions are cognitively represented in terms of their sensory consequences, which might imply that sensory feedback from both one's own and the co-actor's actions is crucial for the SSE to occur. Alternatively, it might be that it is not the other person's action that matters the mere possibility of acting might suffice. If so, an actor should show a SSE even if he or she is unable to perceive the co-actors action and continuously monitor his or her presence.

In an auditory version of the Simon task, Ruys and Aarts (2010) provided actors with relatively constant (online) sensory information about the co-actor's presence by presenting them with colored-light flashes that signaled the co-actor's responses. Even though actors could not see their co-actor, a full-blown SSE was obtained. This outcome demonstrates that it is not the shared presence in the same room that is important for the SSE, but it fails to clarify whether the SSE was due to the sensory feedback about the co-actor's actions or the mere belief that one is collaborating with someone else.

One problem with comparing physical acting with virtual co-acting is that this comparison confounds a number of potentially important factors, such as instructions and the availability of sensory cues. In an attempt to control for the latter, Sebanz et al. (2003) had participants wear earplugs and prevented them from seeing the other person's hand, which did not reduce the SSE. However, the co-actor was still clearly visible as was his/her involvement in the task, which does not render this manipulation particularly strong.

Two recent studies investigated the contributions of online versus offline information more systematically by providing knowledge about a second actor who was said to work on the same task in a different room (Welsh et al., 2007; Tsai et al., 2008). However, while Tsai and colleagues showed clear evidence for effects of offline information (i.e., a SSE was obtained in the physical absence of the co-actor) Welsh and colleagues did not, which renders the evidence equivocal. For evaluating this discrepancy it is informative to consider the set-up of the tasks. Tsai et al. (2008) invited participants who were already acquainted with one another prior to the testing day and allowed them to communicate via intercom before the task and during the break. In contrast, in the study of Welsh et al. (2007) the experimenter was the co-actor, who did not remind the actor of their interaction after having left the room. In other words, the actor's belief that the co-actor would still collaborate with him/her was not updated. It could thus be that offline information about the co-actor is not sufficient to establish the SSE if it is not constantly updated by online information. Therefore it is still not clear what role online information of the co-actor plays in the SSE.

In the present study, we controlled for previous acquaintance with the co-actor and made an attempt to manipulate the availability of sensory feedback about the other in a more systematic fashion. To increase control over perceptual cues, we used an auditory version of the social Simon task. Experiment 1 established this auditory version and was expected to replicate the standard SSE in the auditory domain in accordance with Ruys and Aarts (2010). Experiment 2 included a blindfold condition that eliminated all action-related visual information about the other, to see whether this would reduce or even eliminate the effect. Experiment 3 went one step further by also eliminating auditory cues about the other's actions.

EXPERIMENT 1
METHOD
Subjects
Forty participants (20 male), aged 18–30 (average age: 24.8) were randomly selected from the database of the Max Planck Institute. All participants read and signed an informed consent form for behavioral experiments before being registered into the database. All subjects were right handed (tested according to Oldfield, 1971), had normal or corrected to normal vision and had normal hearing. The subjects were invited as pairs and were asked beforehand if they were already acquainted with one another before the testing day. Acquainted participants could not participate together, and were rescheduled with new co-actors in order to keep *a priori* knowledge of the task and the co-actor as constant as possible for all pairs. Each participant performed a Single Go/No-Go task, a Joint Go/No-Go task (i.e., the Social Simon task) and a standard (solo) Simon task. Each task comprised of the same auditory stimuli. Each participant received 10.50 € for their participation.
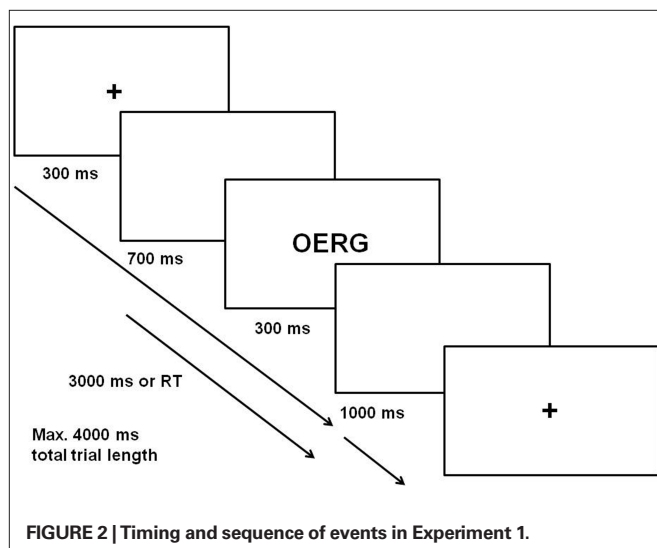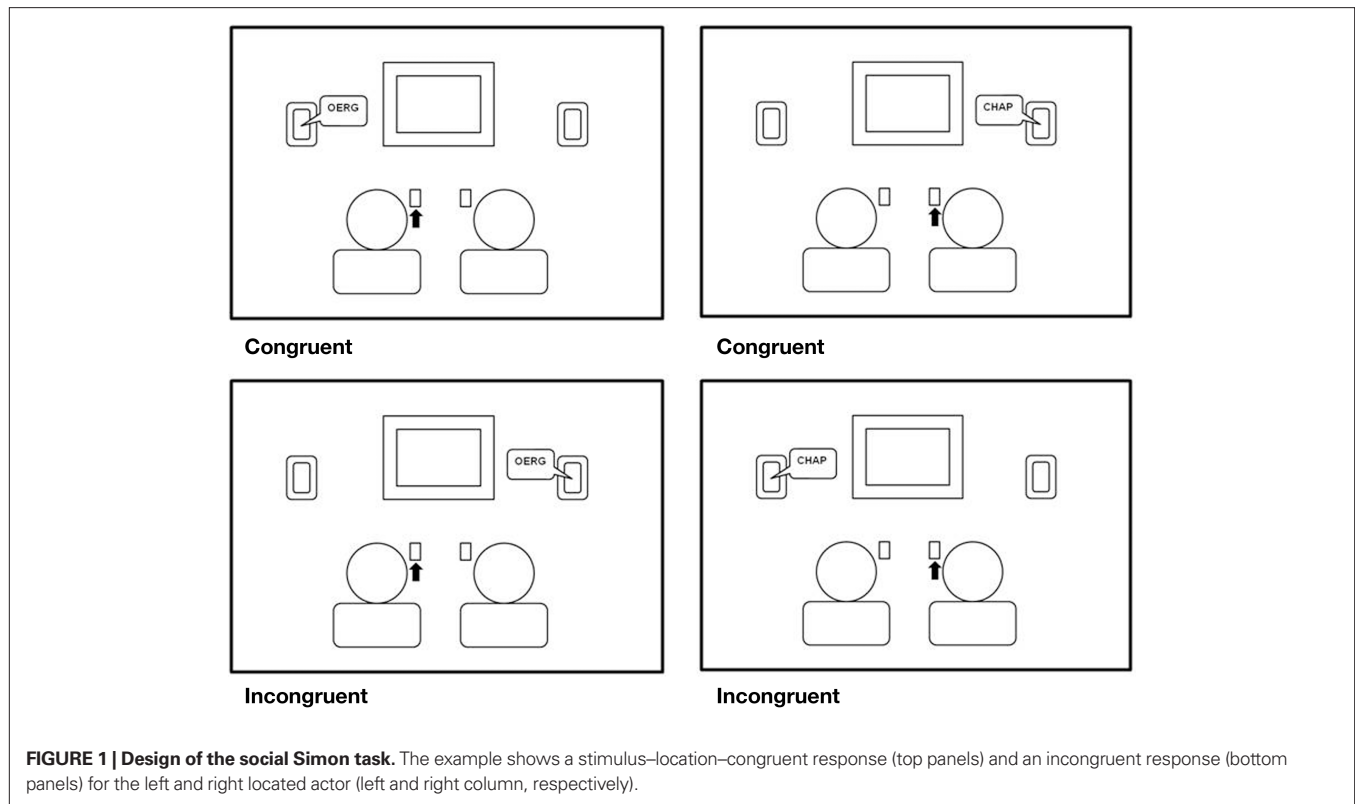
Materials
The auditory stimuli consisted of human vocal utterances without any semantic meaning in German, the testing language. The sounds were originally generated for a functional magnetic resonance imaging study (Henk van Steenbergen, unpublished). We used the reversed and compressed Dutch words "groen" (green) and "paars" (purple) spoken by different male actors and processed using Adobe Audition 2.0 – which resulted in stimuli sounding like "oerg" and "chap." The sounds were adjusted to equal lengths of 300 ms and presented with a loudness of approximately 60 dB. Two loudspeakers were placed 50 cm to the left and right from the middle of a computer screen. Response buttons were placed 25 cm away from the computer screen, 30 cm apart from each other.

Study design and procedure
A 2 (congruent, incongruent) × 2 (Go, No-Go) × 3 (Single, Joint, Standard) factorial design was used. There were 64 trials per design cell for the Single Go/No-Go and the Standard Simon task, and 128 trials per cell for the Joint Go/No-Go task. To keep track of the performance, a feedback screen was presented after half of the trials in each condition. The feedback showed the average reaction times (RTs) and percentage correct (PC), which in the Joint condition referred to the mean performance across both participants. The task was preceded by a training phase of 25 trials per cell.

The two auditory stimuli "oerg" and "chap" were assigned to the left and right button, respectively. In the Joint condition, one participant responded to the "oerg" sound with the left button and was thus seated on the left side while the other participant responded to the "chap" sound and was seated on the right side (see **Figure 1**).

Each trial began with a warning signal, a 300 ms beep presented through both loudspeakers (symbolized by the fixation mark in **Figure 2**). After a silent period of 700 ms, the stimulus tone appeared for 300 ms through the left or right loudspeaker.

FIGURE 1 | Design of the social Simon task. The example shows a stimulus–location–congruent response (top panels) and an incongruent response (bottom panels) for the left and right located actor (left and right column, respectively).

FIGURE 2 | Timing and sequence of events in Experiment 1.

The trial ended after the response was emitted but no later than 3000 ms after stimulus onset. The next trial began after another blank interval of 1000 ms.

The (social) Simon effect was measured by subtracting RTs for incongruent trials (no correspondence of stimulus location and response) from RTs for congruent trials (correspondence of stimulus location and response). Each participant performed the task under three conditions. In the Single condition, participants carried out the task alone in a separate room, sitting on one side and only responding to one sound. In the Joint condition, two participants sat in the same room, side by side, and on the same side as in the Single and Standard conditions. In the Joint condition, each participant responded to the same sound as in the Single condition. In the Standard condition, participants sat in separate rooms and responded to both sounds, but still sat on the same side as in the other two conditions. The order of Single and Joint condition was counterbalanced. The Standard task was presented last as a control condition.

RESULTS

All analyses were tested with an alpha of 0.05. The error rate was very low (Single = 0.5%, Joint = 0.6%, Standard = 3.8%) and error trials were excluded from analyses. The median RTs per participant for correct responses were entered into a two-way repeated measures ANOVA, with Type of task (Single Go/No-Go, Joint Go/No-Go, Standard Simon) and Congruency (congruent, incongruent) as independent factors (for average RTs see **Table 1**). There was a main effect of Congruency ($F(1,39) = 95.72, p < 0.001, \eta^2 = 0.711$); responses were slower in incongruent than in congruent trials ($M = 330$, SE = 9.8, and $M = 312$, SE = 9.0 respectively). The main effect of Type of task was not significant but the interaction between Congruency and Type of task was ($p < 0.001$). Paired-samples tests between congruent and incongruent trials revealed a significant congruency effect in the Standard ($t(39) = 14.48, p < 0.001$) and the Joint condition ($t(39) = 6.04, p < 0.001$), but not in the Single condition ($t(39) = -0.51, p = 0.61$). Given that the Joint condition comprised of twice as many trials as the other two conditions, we re-analyzed the data by considering only the first 64 trials per cell of the Joint condition, but the outcome was the same.

Table 1 | Experiment 1, mean RTs and Simon effect for the three conditions.

	Congruent	Incongruent	Simon effect
Single	323	321	–2
Joint	302	315	13
Standard	311	353	42

DISCUSSION

The outcome of Experiment 1 is straightforward: a Simon effect was obtained both in the standard and in the joint-action condition but not in the single condition. This replicates the basic findings of Sebanz et al. (2003) and extends it to auditory stimuli (in accordance with Ruys and Aarts, 2010).

EXPERIMENT 2

The aim of Experiment 2 was to eliminate visual action-related information about the co-actor without changing any other aspect of the experimental task, the context, and the instruction. We did that by having all participants wear goggles that in one group of participants were translucent, which would basically put them into the same situation as the participants of Experiment 1, but that in another group of participants were opaque. Thus, in this group, no visual online information was available, even though the participants were aware of the presence of their co-actor and heard him/her carry out the task. If visual online information would be relevant for the participant's continuous grounding of the task representation, the SSE should be weaker or absent in the blindfolded group. Alternatively, if *a priori* knowledge (offline information) is sufficient to establish the SSE, while online information is merely redundant, then we should find no reduction of the SSE in the blindfolded group.

METHOD

Forty-two participants (18 male), aged 18- to 30-years old (average age: 23.19), were selected according to the same criteria applied in Experiment 1. Each participant received 7.50 €. One pair of subjects violated the instructions not to talk during the experiment and their data were removed from analyses. The method was the same as in Experiment 1, with the following exceptions. Participants in the seeing group wore see-through glasses while participants in the blindfolded group wore opaque glasses (see **Figure 3**).

A Joint Go/No-Go task similar to the Joint condition of Experiment 1 was used. The task employed a 2 (Go/No-Go) × 2 (congruent, incongruent) × 2 (visual information present or absent) factorial design. Participants were presented with a feedback screen after half of the trials, blindfolded subjects were allowed to take off their goggles to see it. The task consisted of 128 trials per cell for each participant (in total 512 trials were presented). It was preceded by a training phase of 25 trials per cell (during the training phase the participants in the blindfolded condition were already blindfolded). Participants were instructed not to talk to each other during the experiment.

RESULTS

The error rate was again very low (1.2%). Median RTs for correct responses were entered into a two-way mixed ANOVA, with the independent variable Congruency (congruent, incongruent)

FIGURE 3 | Goggles used in the blindfolded (see left participant) and seeing conditions (see right participant).

as within-subjects factor and the independent variable Visual Feedback (present, absent) as between-subjects factor. There was a main effect of Congruency ($F_{(1,38)} = 122.55$, $p < 0.001$, $\eta^2 = 0.76$); responses were slower in incongruent than in congruent trials ($M = 306$, SE = 7.9, and $M = 284$, SE = 7.2 respectively). There was neither a main effect of Visual Feedback ($F_{(1,38)} = 0.07$, $p = 0.8$, $\eta^2 = 0.002$), nor a significant interaction ($F_{(1,38)} = 1.04$, $p = 0.314$, $\eta^2 = 0.03$), suggesting that the Simon effects were equivalent in the two conditions (**Table 2**).

DISCUSSION

There was no evidence whatsoever that eliminating visual online feedback about the co-actor reduced or eliminated the SSE – the numerical effect was even larger in the absence of visual information. Given that participants were blindfolded even during the training phase, each participant had only very little information about the co-actor's actions to improve on that during the task. This suggests that action and task representations do not rely on online information, but on *a priori* knowledge (offline information) to interact with a social, intentional interaction partner. However, in Experiment 2 auditory action-related online information from the button presses may have established the SSE in the blindfold condition, an issue that we addressed in Experiment 3.

EXPERIMENT 3

Although participants in Experiment 2 were prevented from processing visual online feedback, they did have access to auditory online feedback. Both co-actors were using buttons of a standard keyboard, which provided sensory cues about the other's continuous presence and responses. Experiment 3 aimed to assess the contribution from this auditory information by having pairs of seeing and blindfolded participants working either with a standard keyboard that did provide auditory feedback or with a noise-free keyboard that did not. If online auditory action-related feedback from the co-actor would play a role, the SSE should be reduced or disappear with a noise-free keyboard. Alternatively, if *a priori*

Table 2 | Experiment 2, mean RTs and Simon effect for the two conditions.

Visual feedback	Congruent	Incongruent	Simon effect
Present	283	303	20
Absent	285	309	24

Table 3 | Experiment 3, mean RTs and Simon effects as a function of visual and auditory feedback.

Auditory feedback	Visual feedback	Congruent	Incongruent	Simon effect
Present	Present	325	350	25
	Absent	374	382	8
Absent	Present	342	362	20
	Absent	379	405	26

knowledge (offline information) is sufficient to establish the SSE, then we should find no reduction of the SSE when eliminating visual and auditory online information.

METHOD

Forty participants (18 male), aged 18- to 30-years old (average age: 23.14), were selected according to the same criteria as in Experiment 1. The method was as in Experiment 2, with the following exceptions. In addition to the manipulation of the visual feedback between participants, the presence of auditory feedback (present, absent) was manipulated within participants. Each participant performed one block with a standard keyboard and another block with a noise-free keyboard, with the order being balanced across participants. To shorten the experiment, the length of each trial was reduced to a maximum of 2000 ms. Each participant worked through 32 trials per cell, 256 trials in total. The task was preceded by a training phase of eight trials, two per cell, during which the participants in the blindfold condition were again already blindfolded.

RESULTS

The error rate was again very low (0.8%). Median RTs for correct responses were entered into a three-way mixed ANOVA, with Congruency (Congruent, Incongruent) and Auditory Feedback (present, absent) as within-subjects factors and Visual Feedback (present, absent) as between-subjects factor. There was a main effect of Congruency ($F(1,38) = 40.99$, $p < 0.001$, $\eta^2 = 0.52$); Responses were slower in incongruent than in congruent trials ($M = 375$, $SE = 12.5$, and $M = 355$, $SE = 11.7$ respectively). There was neither a main effect of Visual Feedback ($F(1,38) = 2.84$, $p = 0.1$, $\eta^2 = 0.07$), nor of Auditory Feedback ($F(1,38) = 2.52$, $p = 0.12$, $\eta^2 = 0.06$), nor any significant interaction. The average Simon effect was similar for participants with both auditory and visual feedback (25 ms) and the participants without any feedback (26 ms; **Table 3**).

DISCUSSION

Despite having no online feedback about the other's actions, participants showed a full-blown SSE and there was not even a sign of a reduction of the effect in the absence of visual and auditory feedback. The only peculiarity in the numerical data pattern is the rather small effect in the condition with auditory but without visual feedback. However, given that, in Experiment 2, the same condition yielded a full-blown SSE comparable to the other conditions in Experiment 3, we consider this an accidental observation of no theoretical relevance. In any case, it seems clear that online visual or auditory feedback from the other is not required for the SSE to occur. Instead, the present findings suggest a central role of *a priori* knowledge (offline information) and the belief to interact with a social, intentional agent (Tsai and Brass, 2007) for the SSE.

CONCLUSIONS

The main aim of our study was to investigate the contribution of online visual and auditory information about a co-actor to the SSE. The very existence of the SSE suggests that action and task representations are grounded in the current situational context and consider cues about the presence and activities of co-actors. However, our present findings suggest that this grounding does not need to be continuous, in the sense that these representations can survive in the absence of ongoing visual and auditory feedback. After having established our auditory version of the social Simon task in Experiment 1 and replicated the basic findings reported by Sebanz et al. (2003), we tested the contribution of visual feedback from the other in Experiment 2 and the contribution of auditory feedback about the other's actions in Experiment 3. Even though our manipulation of auditory feedback does not rule out task-unrelated feedback from the co-actor, such as breathing noises or coughs, participants in the no-visual/no-auditory condition of Experiment 3 did not have any sensory cues about the action being performed by the other. And yet, a full-blown SSE was obtained.

What matters for the SSE does not seem to be online information about the social situation but the mere knowledge that a social, intentional co-actor is present. This conclusion does not support the concept of co-representation suggested by Sebanz et al. (2003). If the SSE would emerge at a representational level that does not distinguish between one's own actions and the actions of another person, and if that representational level would be fed by sensory feedback about both types of actions, one would expect the SSE to strongly rely on more or less continuous sensory action feedback. Eliminating this feedback should thus eliminate the SSE, which is not what our present findings show. Instead, what seems to matter is apparently the actor's belief that he/she is interacting with an intentional agent (Tsai and Brass, 2007), which is likely to rely on *a priori* knowledge about the intentional co-actor.

Hence, top-down effects (Liepelt and Brass, 2010) seem to be much more central to the SSE than previously thought. Top-down modulation may be even more important in the SSE than, for example, in automatic imitation research, where taking away the actor's intention reduces but does not eliminate stimulus–response priming (Liepelt et al., 2008).

This is likely to explain why the SSE is eliminated if the actor is led to believe to interact with an un-intentional agent (Tsai and Brass, 2007). It also provides some pointers to why Tsai et al. (2008) were able to produce an SSE but Welsh et al. (2007) were not. As discussed already, the participants of Tsai et al., but not

of Welsh et al., were repeatedly updated about the presence of their co-actor, which was likely to strengthen the actor's belief in this presence.

To summarize, we show that task representations can be grounded in offline information, but the contradictory findings between Welsh et al. (2007) and Tsai et al. (2008) remind us that this offline information may be kept active for only a limited amount of time. Note that our setup did not prevent participants from seeing and talking to each other upon arrival, before the actual experiment began, and it might be that experience (or the memory of it)

that provides the sensory grounding for the task co-representations responsible for the SSE. In any case, we can conclude that task co-representations do not rely on online feedback about the other person's actions.

REFERENCES

Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representation*. New York: Cambridge University Press.

Elsner, B., and Hommel, B. (2001). Effect anticipation and action control. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 229–240.

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychol. Res.* 73, 512–526.

Hommel, B., Colzato, L. S., and Wery van den Wildenberg, P. M. (2009). How social are task representations? *Psychol. Sci.* 3, 1–5.

James, W. (1890). *The Principles of Psychology*, Vol. 2. New York: Dover Publications.

Liepelt, R., and Brass, M. (2010). Top-down modulation of motor priming by belief about animacy. *Exp. Psychol.* 57, 221–227.

Liepelt, R., von Cramon, D. Y., and Brass, M. (2008). What is matched in direct matching? Intention attribution modulates motor priming. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 578–591.

Lotze, R. H. (1852). *Medicinische Psychologie oder die Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Ruys, K. I., and Aarts, H. (2010). When competition merges people's behavior: interdependency activates shared action representations. *J. Exp. Soc. Psychol.* 46, 1130–1133.

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21.

Simon, J. R., and Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *J. Appl. Psychol.* 51, 300–304.

Stock, A., and Stock, C. (2004). A short history of ideo-motor action. *Psychol. Res.* 68, 176–188.

Tsai, C.-C., and Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Co-acting with a human hand versus a wooden hand in a dyadic interaction. *Psychol. Sci.* 18, 1058–1062.

Tsai, C. C., Kuo, W. J., Hung, D., and Tzeng, O. (2008). Action co-representation is tuned to other humans. *J. Cogn. Neurosci.* 20, 2015–2024.

Welsh, T. N., Higgins, L., Ray, M., and Weeks, D. J. (2007). Seeing vs. believing: is believing sufficient to activate the processes of response co-representation? *Hum. Mov. Sci.* 26, 853–866.