

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

The implicit effect of action mental simulation on action evaluation

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1850580> since 2024-03-22T14:05:30Z

Published version:

DOI:10.1177/17470218221091096

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

The implicit effect of action mental simulation on action evaluation

Francesco Iani^{ab}, Teresa Limata^a, Monica Bucciarelli^{ab}, Giuliana Mazzoni^c

^a Dipartimento di Psicologia, Università di Torino

^b Centro di Logica, Linguaggio, e Cognizione, Università di Torino

^c Dipartimento di Psicologia, Sapienza Università di Roma

Corresponding author:

Teresa Limata

Università di Torino

Dipartimento di Psicologia

Via Verdi, 10

10124 Turin, Italy

e-mail: teresa.limata@unito.it

tel: +39.011.6703038

Abstract

The main assumption underlying the present investigation is that action observation elicits a mandatory mental simulation representing the action forward in time. In Experiment 1 participants observed pairs of photos portraying the initial and the final still frames of an action video; then they observed a photo depicting the very same action but either forward or backward in time. Their task was to tell whether the action in the photo portrayed something happened before or after the action seen at encoding. In this explicit task the evaluation was faster for forward photos than for backward photos. Crucially, the effect was replicated when instructions asked only to evaluate at test whether the photo depicted a scene congruent with the action seen at encoding (implicit task from two still frames, Experiment 2), and when at encoding they were presented a single still frame and evaluated at test whether a photo depicted a scene congruent with the action seen at encoding (implicit task from single still frame; Experiment 3). Overall, the results speak in favour of a mandatory mechanism through which our brain simulates the action also in tasks that do not explicitly require action simulation.

Keywords: Anticipation, Mental simulation, Action perception, Mandatory simulation, Action photos

Words count: 10735

Public Significance Statements

Former studies have revealed that when people observe others' actions their motor brain areas become active as if they were performing the very same actions.

This finding is well established in the neuroscientific literature, but difficulties arise in delineating and defining the exact nature of the mental processes associated to these activations.

The results of three experiments provide evidence for two important features of the cognitive processes underlying action observation.

Action observation triggers a simulation of the action unfolding in time, thereby representing something that was not actually perceived.

This process is mandatory, as it occurs independently of the will of the observer.

Overall considered, these results speak in favour of the existence of a mandatory mechanism through which our brain mentally anticipates the observed action.

1. Introduction

The observation of others' actions involves the activation of the same neuronal areas responsible for action production, in both non-human and human primates (e.g., Gallese et al., 1996; Buccino et al., 2001). An ever-growing literature has detected the activation of a neural network that mainly involves the premotor cortex, the inferior frontal cortex, the inferior parietal lobule and the supplementary motor area (the so-called "action observation network" - AON; for a meta-analysis see, e.g., Caspers et al., 2010). The finding is well established in the neuroscientific literature, but difficulties arise when interpreting these activations in terms of their functional role and their behavioural consequences. Among the different theses, these processes would be fundamental in understanding others' intentions (see, e.g., Gallese, 2007), they would constitute the first step in reaching a complete mental representation of others' goal (Goldman, 2009), they would be responsible only for low-level processing (for a review, see, Heyes & Catmur, 2020), or they would simply reflect a Pavlovian association (i.e., a sensory-motor pairing, Mahon & Caramazza, 2005). The great deal of uncertainty about the function of the AON is probably due to the difficulty to delineate and define the mental processes associated to these activations, to such an extent that the same term "simulation" has been used in reference to quite different processes involved in quite different cognitive phenomena such as language comprehension (Fischer & Zwaan, 2008), memory retrieval (Kent & Lamberts, 2008) and spatial mental imagery (Moulton & Kosslyn, 2009). And indeed, some scholars have highlighted the importance to discriminate the different nature of mental simulations underlying different cognitive phenomena, as well as the different behavioural consequences they feature (see, e.g., Ianì et al., 2019).

The focus of the present investigation is mental simulation stemming from action observation. Specifically, it aims at gathering more clear evidence about the nature of mental processes underlying the observation of other people's actions. A possible theoretical framework suggests that cognitive processes involved in action observation might operate as other predictive mechanisms of our cognitive system that allow us to effectively interact with the environment (e.g., Donnarumma

et al., 2017). Prediction and simulation are fundamental embodied functions of the cognitive system: we are extraordinary machines that travel forward in time and we anticipate what is going to happen in several domains. The nervous system generates predictions by running “online” simulations (Wilson, 2002): as the external event unfolds, the simulated event likewise unfolds in the same way and at the same pace. These mental representations of a sequence of events as they unfold in time are akin to the idea of perceptual emulators (Grush, 2004; see also Wilson & Knoblich, 2005). The idea is that the brain constructs neural circuits that act as models of an external system that run simulations of that system in real time. We know from the literature that humans normally run mental simulations forward in order to anticipate the trajectory of moving objects (Pozzo et al., 2006), as well as to anticipate linguistic inputs, an ability that facilitates fluent reading and writing acquisition (see, e.g., Guasti et al., 2017; Järvillehto et al., 2009). In this view, motor reactivity to observed actions likely reflects the anticipatory simulation of future phases of the observed action (e.g., Schütz-Bosbach & Prinz, 2007).

Within this proposed theoretical framework, action observation elicits the construction of a kinematic mental simulation of the action unfolding in time (i.e., a forward simulation) that provides both sensorimotor information about the perceived action, and sensorimotor information about proximal states of the action (Iani et al., 2018). A key feature of this kind of mental simulation is that it may allow an observer to “covertly” anticipate others’ behaviours. Consistent with this assumption, experimental evidence suggested that the attribution of a given intention to an actor relies on a mental simulation guided by the kinematics of the observed action (which in turn depends on the actor’s intention; see, for a review, Quesque & Coello, 2015), as well as by the objects in the peripersonal space of the observed agent (e.g., Cardellicchio et al., 2013). In other words, individuals use implicit procedural motor knowledge in predicting the unfolding in time of a perceived action. In line with this assumption, false memories occur for actions that represent the unfolding over time of the static action initially observed (Iani et al., 2020).

Based on evidence that forward simulation relies on the sensory-motor system and it is guided by action's kinematic features, we assumed that it is somehow a process independent of cognitive control and deliberative reasoning (see also Gallese, 2007). Hereafter we call this process a 'mandatory forward simulation', by which we simply mean a forward simulation of the action that takes place even when there is no explicit request and no need to consider the direction of the action.

2. On the consequences of forward mental simulation

Our former studies sought evidence of the existence of a forward simulation stemming from action observation (Ianì et al., 2021). In a first experiment, the task of the participants was to observe videos, each depicting the central part of an action performed by an actress on an object (e.g., eating a hamburger). Soon after each video, participants were shown a photo, either depicting the initial part or the final part of the whole action, i.e., states of the action not observed in the video. Their task was to evaluate whether the photo portrayed something happened before (backward photo) or after the action depicted in the video (forward photo). As the results revealed, evaluation of forward photos was faster than evaluation of backward photos, thus suggesting that forward simulation is most likely to be very rapid and mandatory. In a second experiment we explored how much forward simulations rely on the observer's motor resources. The results showed that keeping the hands crossed behind the back interfered with forward simulations, that is participants were not faster in recognizing forward compared to backward frames. In other words, when participants kept their arms crossed behind their back, they struggled in simulating the ongoing action (i.e. they did not show the advantage in evaluating forward photos). We interpreted this result as indicating that mental simulations from action observation rely at least in part on the observers' potentiality for action, as bodily conditions incompatible with the observed action hinder the tendency to recognise forward photos faster than backward ones. This means that observers

“exploit” their body to simulate the observed action: if their body is less available (as in the crossed hand condition) also their cognitive mechanisms are affected (Iani et al., 2021).

In Experiment 1, we sought to provide more evidence for the forward effect discovered in our earlier study (Iani et al., 2021) by improving the original experimental paradigm. Further, Experiments 2 and 3 aim to investigate unexplored consequences of action prediction by attempting to answer two different questions: (1) “Is the simulative mechanism merely a resonance of what is seen, or does its main aim consist in representing the ongoing action (i.e. something not yet observed)?”; and (2) “Can we regard this simulative mechanism a mandatory process which takes place also without explicit demands?”. Thus, the present investigation aims to gather clearer evidence that cognitive mechanisms stemming from action observation are mandatory simulations of what is going to happen.

In our previous studies, for each video seen at encoding, at test participants saw either a backward or a forward photo. Since the first stimulus was a video, the temporal distance between the presentation of the first frame of the video and the backward photo was greater compared to the distance between the presentation of the last frame of the video and the forward photo. The difference in temporal distances could represent an alternative explanation of the reported forward effect¹. In Experiment 1 of the present investigation, we aimed to control for the role of this potential artefact in results. At encoding, participants saw both the first and the last frame of the central part of the original videos (videos here not presented), rather than watching the video itself, thereby observing at the same time the onset and the offset of the central part of the action. Thus, the temporal distance between the presentation of the first frame of the video and the backward photo was the same as the temporal distance between the presentation of the last frame of the video and the forward photo.

¹ However, it should be noted that in Iani et al. (2021) the first frame of these videos remained on the computer screen for 1.000 ms before the actual video presentation (whose average length was 2.060 ms); this procedure was meant to partially counterbalance the difference in temporal distance between backward and forward photos, on one side, and the initial and the final part of the video.

Furthermore, in our previous studies (and in the current Experiment 1) participants were explicitly invited to consider whether the test photo represented something that happened before or after what was portrayed in the relative video (in the current Experiment 1 the two photos), an instruction that could elicit deliberative mental simulation rather than an implicit mandatory mental simulation. To address this issue and ensure to rule out the role of a deliberate, intentional mental simulation, participants in Experiment 2 of the present investigation were just invited to consider whether the photo at test depicted a scene congruent with the action seen in the photo pairs. The term ‘congruence’ referred to whether the general action in the two photos was the same or whether it was different (i.e. there was no reference to the direction of the action).

However, seeing at encoding a pair of photos might lead participants to have first a look at the onset photo and only afterwards, temporally distanced, at the offset photo. To exclude the possibility that results could be due to this artefact, in Experiment 3 participants observed at encoding just a single photo of the action, and at test they were invited to consider whether a new unseen photo depicted a scene congruent with the action seen in the previous photo. Furthermore, by presenting just one photo one rules out the possibility to implicitly suggest the participant to fill in the gap between the onset photo and the offset photo; if by presenting just one photo one obtains the same results (faster forward evaluations of congruency) then one can conclude that a forward mental simulation occurs anytime one sees an action, even when frozen in time, as in a single still frame. This might provide stronger evidence in favour of the thesis that seeing an action mandatorily triggers a forward kinematic mental simulation of the action.

The Bioethical Committee of Turin University approved the investigation.

3. Experiment 1: Evaluation of forward photos is faster than evaluation of backward photos

The participants in Experiment 1 observed pairs of photos each depicting the first and the last frame of the central part of an action. After observing each pair, at test participants saw either a backward or a forward photo. Backward photos portrayed something happened before what was

represented in the first photo seen at encoding (i.e., the central frame of the first part of the action), whereas forward photos portrayed something that happened after what was represented in the second photo of the pair (i.e., the central frame of the last part of the action). Their task was to say “before” or “after” depending on whether they thought the action in the photo portrayed something that happened before or after the action portrayed in the photos pair, respectively. We used vocal responses in place of manual responses to not load up the same hand-related motor resources required to process the observed action. The prediction was that evaluation of forward photos should be faster than evaluation of backward photos.

3.1. Method

3.1.1. Participants

Thirty-six students from Turin University (14 males and 22 females, mean age = 24.8, $SD = 5.4$) voluntarily took part in the experiment in exchange of a course credit. They previously signed the informed consent. The sample size for all experiments was determined through a simulation-based power analyses (Kumle et al., 2021) using the reaction times data obtained in a pilot research (Ianì et al., 2021). Based on the mixed-effect model including the same fixed and random factors of those reported in all experiments here, 1000 new data sets, each containing n participants (16,26,36,46), were simulated using the `mixedpower()` function in R. For $n = 36$ simulated participants, we estimated a statistical power of .803, that is for 803 out of 1000 simulation, the model detected a significant effect. Then, for each experiment we tested at least 36 participants.

3.1.2. Material

The material consisted in 10 videos from our former study (Ianì et al., 2021). Each video depicts a single action performed by an actress with the upper limbs, with either one or two arms (e.g., answering the phone; see the description of each test in Table 1; see the full set in the Supplementary Materials). Each video was cut into three same-length parts and we extracted from the middle part the initial and the last frame. These two photos were presented to the participants at

encoding. The material included the middle frame of the first/initial part of the whole video (backward photo) and the middle frame of the third/final part of the whole video (forward photo). These two photos were presented at test. In Figure 1 are reported examples of the stimuli obtained for the video “Eat a hamburger”.

Table 1

The description of the action performed by the actress in each scenario.

Number of scenarios	Description
1.	Drink a glass of water
2.	Open a book
3.	Use eye drops
4.	Close a portable computer
5.	Put on a hat
6.	Eat a hamburger
7.	Peel a banana
8.	Tear a sheet of paper
9.	Sniff a flower
10.	Blow the nose

Figure 1

The pairs of photos extracted from the video “Eat a hamburger” and presented contemporaneously to the participants at encoding (First photo and Second photo) and the photos presented at test (Backward photo and Forward photo).



Left photo (encoding)



Backward photo (test)

Right photo (encoding)



Forward photo (test)

3.1.3. Design and procedure

We created two experimental protocols. Referring to the stimuli described in Table 1, in Protocol 1 stimuli from 1 to 5 were paired with backward photos and stimuli from 6 to 10 were paired with forward photos, in Protocol 2 stimuli from 1 to 5 were paired with forward photos, and stimuli from 6 to 10 were paired with backward photos. Half of the participants were assigned to Protocol 1 and half to Protocol 2. The presentation order of the pairs of photos (and relative test photos) was randomized for each participant in each protocol using E-Prime 3.0 Software.

The experiment took place online. Participants received an email with the instructions to download E-prime Go, an extension of the software to run online experiments. In the first screenshot, participants received the following instructions:

Thank you for participating in this experiment on how we understand the unfolding of an action over time. You'll see pairs of photos that represent the same action in different moments. In each pair, the right-side photo depicts a later moment of the action. Soon after each photos pair a cross of fixation will appear, followed by another photo. Your task is to decide whether the photo depicts what happened before or after the action seen in the photos. Answer out loud by saying BEFORE if you think it happened before, by saying AFTER if you think it happened after.

When ready to start the participant pressed the space bar and the first pair of photos appeared on the screen for 3 seconds, followed by a cross of fixation lasting 250 ms. Then, the test photo appeared

on the screen for 7 seconds. This inter-stimulus interval (ISI) granted that participants had to rely on simulation mechanisms; indeed, ISIs shorter than 150 ms generate apparent visual motion that relies on perceptual mechanisms (Verfaillie & Daems, 2002). The test photo was followed by a white screen with the text “Next pair” lasting 3 seconds.

We measured participants’ response times with MatLab 2020 software, from photo presentation to participants’ oral answer. Specifically, RTs were measured by calculating milliseconds from photo presentation onset to oral answer onset (the MatLab script used to calculate RTs is available online in the Supplementary Materials). We chose this solution, compared to measuring milliseconds from photo presentation to oral answer ending, to avoid potential differences in the lexical processing required to say “before”/ “after” (“prima”/ “dopo” in Italian).

3.2. Results

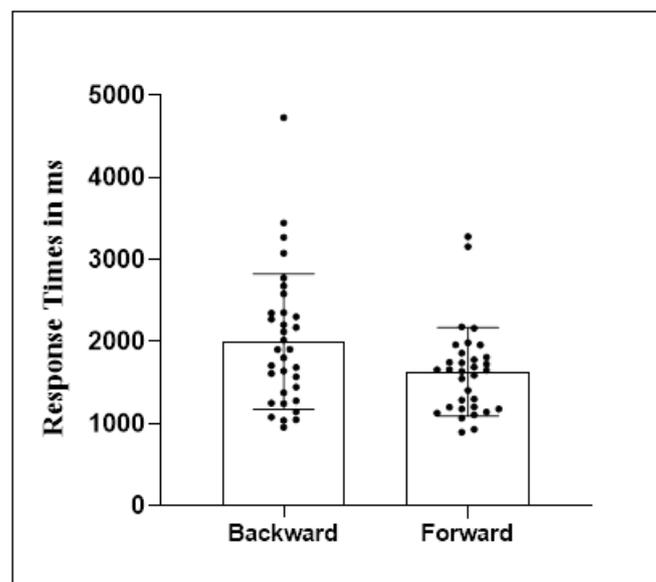
The data of 3 of the 36 participants were dropped from the analysis because the audio recording of two of them was disturbed and one was an outlier. Outliers were defined as values higher than (3rd quartile + 1.5*interquartile range (IQR)) or lower than (1st quartile – 1.5*IQR); the participant’s accuracy with forward photos fell below the lower limit. For the remaining 33 participants we verified whether the type of test photo (forward vs backward) influenced participants’ reaction times (log-transformed) for correct responses using a linear mixed effect regression model implemented with the `lmer()` function from the `lmerTest` package (Kuznetsova et al., 2017) in the R statistical programming environment (version 4.1.0; R Development Core Team, 2021). The same package was used in order to estimate p values and the model.

Following the guidelines in the psycholinguistic literature (Barr et al., 2013), we started from the theoretical full model (by including the maximal structure of random effects supported by the design) and then applied to this model the *step function* in the `lmerTest` package (Kuznetsova et al. 2017) to find the final mixed model in which the factors with non-significant effects were removed. This function performs backward elimination of non-significant effects (both random and fixed). Appendix B gives the theoretical full model for each experiment and the final model found by the

step function in the package lmerTest. The final model revealed a significant effect of Type of photo ($\beta = .067$, $SE = .018$, $t = 3.77$, $p < .001$) on log-transformed reaction times. Participants were faster in correctly evaluating forward photos (a mean of 1630 ms, $SD = 539$ ms) compared to backward photos (a mean of 2001 ms, $SD = 825$ ms). Figure 2 shows RTs for correct evaluations in backward and forward conditions. The RTs dispersion in the two experimental conditions in Figure 2 suggests that participants' assessment was more uncertain in backward condition than in forward condition; this is exactly what should occur if the backward photos do not match with the action mental simulation resulting from the forward mental simulation.

Figure 2

Response times (in ms) and standard deviations for correct evaluations in Experiment 1.



We also analysed the likelihood of error using a mixed-effect logistic regression, implemented with the `glmer()` function from the `lme4` package (version 1.1-26). The model random effect structure included random intercepts for both participants and items ($\text{Accuracy} \sim \text{Type of photo} + (1|\text{Subjects}) + (1|\text{Items})$). The results revealed a significant effect of Type of photo on likelihood of error ($\beta = 1.43$, $SE = .41$, $z = 3.48$, $p < .001$). Participants' likelihood of error was greater for

backward photos compared to forward photos (a mean of .81, $SD = .22$; a mean of .94, $SD = .09$, respectively).

Table A1 in the Appendix summarizes means and standard deviations for response times and errors by the participants in the experiment.

3.2.1. Discussion

The results of Experiment 1 showed that still frames of actions taken forward in time compared to what has been actually observed are faster to evaluate compared to still frames taken backward in time. This shows that the greater speed of evaluation of forward photos observed in previous studies (Ianì et al., 2021) did not depend on an artefact due to time proximity between the last part of the action actually seen and the forward action to be evaluated.

However, one can claim that in this experiment participants intentionally ran a mental simulation of the action because of the explicit request to state whether the still frames at test were depicting forward or backward moments of the action. Thus, it might be the case that forward simulation occurs only when instructions trigger a *voluntary* forward mental simulation. The idea that forward mental simulation is mandatory and independent from the instructions received would be further strengthened in case individuals were found to be faster in evaluating forward photos compared to backward photos also in the absence of any explicit demand to evaluate their temporal relation with the pair of photos encountered at encoding. This is addressed in Experiment 2.

4. Experiment 2: Implicit evaluation is faster for forward than backward photos (implicit task after observation of a pair of photos)

The experiment was a replication of Experiment 1 except for the instructions given to the participants. Differently from Experiment 1, to avoid the intentional running of a mental simulation of the actions determined by the requests in the instructions, in Experiment 2 participants were simply invited to decide whether each photo at test depicted a scene that was congruent (or not) with the action observed in the pair of photos seen at encoding. We clarified to participants that

“congruent” referred to whether the test photo represented the same action or a different action than the one in the stimulus photo, and participants were given examples (e.g., eating a hamburger is different from opening a book). Thus, there was no reference to the direction of the action, and it was clear that incongruency referred to completely different actions.

The prediction was that, if mental simulation is mandatory, evaluation of forward photos should be faster than evaluation of backward photos also in absence of the explicit demand to evaluate their temporal relation with the pair of photos at encoding.

4.1. Method

4.1.1. Participants

Forty-two students from Turin University (18 males and 24 females, mean age = 25.9, $SD = 5.5$) voluntarily took part in the experiment in exchange of a course credit.

4.1.2. Material

The material was the same as in Experiment 1 (see Table 1).

4.1.3. Design and procedure

Design and procedure were the same as in Experiment 1, with the exception that soon after the presentation of each pair of photos participants saw either a backward or a forward photo and their task was to evaluate whether the photo was congruent with the scenarios observed in the pair of photos. Furthermore, to allow for the evaluation of incongruency of the test photos with the pair of photos seen at encoding, at test for 4 scenarios participants were presented with photos depicting actions that were different from those portrayed in the pairs at encoding (filler stimuli, requiring “NO” responses), for 6 scenarios they were presented with test photos depicting the forward states of the actions portrayed in the pairs at encoding (3 scenarios), and test photos depicting backward actions (3 scenarios, both sets requiring “YES” responses).

We created three experimental protocols, by randomly assigning the video stimuli to three blocks. Referring to the stimuli described and numbered in Table 1, the first block contained stimuli

4, 5, 8; the second block contained stimuli 2, 6, 9; the third block contained stimuli 3, 7, 10. In each protocol, one block of stimuli was used as filler; we used also the stimulus numbered 1 as filler in all protocols, for a total of 4 filler stimuli. In Protocol 1, the first block was paired with backward photos and the second block with forward photos; in Protocol 2, the second block was paired with backward photos and the third block with forward photos; in Protocol 3, the third block was paired with backward photos and the first block with forward photos. Participants were randomly assigned to the three protocols. The presentation order of the pairs of photos (and relative test photos) was randomized for each participant in each protocol.

The instructions were as follows:

Thank you for participating in this experiment on how we understand the unfolding of an action over time. You'll see pairs of photos that represent the same action in different moments. In each pair, the right-side photo depicts a later moment of the action. Soon after each pair of photos is presented, a central fixation cross will appear, followed by another photo. Your task is to decide whether the photo depicts a scene congruent or not with the action seen in the original photos. Answer aloud by saying YES if you think it is congruent, by saying NO if you think it is not.

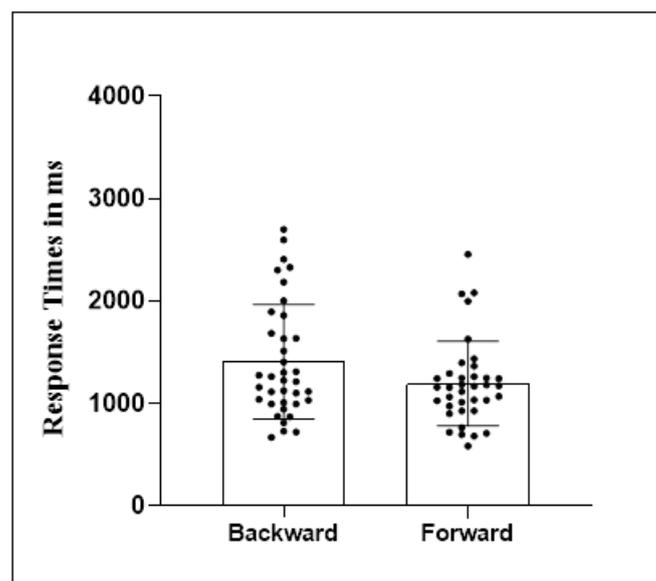
4.2. Results

The results of 5 of the 42 participants were dropped from the analysis because the audio recording of one participant was disturbed and four participants were outliers. The accuracy of three participants with backward photos and the accuracy of one participant with forward photos felt below the lower limit. For the remaining 37 participants we verified whether the Type of photo test (forward vs backward) influenced participants' reaction times (log-transformed) for correct responses using a linear mixed effect regression model using the same R package used in Experiment 1. Following again the guidelines in the psycholinguistic literature (Barr et al., 2013), we started from the theoretical full model (by including the maximal structure of random effects supported by the design) and then applied to this model the *step function* in the lmerTest package

(Kuznetsova et al., 2017) to find the final mixed model in which the factors with non-significant effects were removed (see Appendix B, for the theoretical full model and the final model). The final model revealed a significant effect of Type of photo ($\beta = .055$, $SE = .015$, $t = 3.60$, $p < .001$) on log-transformed reaction times. This means that participants were faster in correctly responding to forward photos (a mean of 1193 ms, $SD = 411$) compared to backward photos (a mean of 1405 ms, $SD = 559$). Figure 3 shows RTs for correct evaluations in backward and forward conditions. As we observed in Experiment 1, also in Experiment 2 the RTs dispersion in the two experimental conditions suggests that participants' assessment was more uncertain in backward condition than in forward condition.

Figure 3

Response times (in ms) and standard deviations for correct evaluations in Experiment 2.



Also, participants' likelihood of error for forward photos and backward photos did significantly differ across backward and forward conditions. Since linear mixed models always generated a singular fit, we ran a simple linear regression model using `lm()` function in R. We did detect a significant effect of Type of photo ($\beta = .16$, $SE = .06$, $t = 2.64$, $p = .01$), suggesting that

participants' likelihood of error was greater in the evaluation of congruent backward photos than forward photos (a mean of .95, $SD = .13$ and a mean of 1, $SD = 0$, respectively).

Table A2 in the Appendix summarizes means and standard deviations for response times and errors by the participants in the experiment.

4.2.1. Discussion

In this experiment, differently from Experiment 1, participants were simply requested to provide an evaluation of congruency (or not) of new still frames of actions (taken forward and backward in time) with seen actions. Actions seen were just two still frames, one depicting the onset of the central part of an action and the other the offset of the same central part of the action. If people were spontaneously running a forward kinematic mental simulation of the action seen, this time without any instructions suggesting doing so, we were expecting the congruency of forward still frames to be evaluated faster than that of backward frames. The results confirmed this prediction, providing then further evidence in favour of the thesis that forward mental simulation occurs spontaneously and is a mandatory process that is triggered by seeing still frames of an action. Our results show that forward mental simulation occurs even when it is detrimental to performance. We are entitled to assume that participants, when invited to assess whether each frame depicted something happened either before or after the critical frame, were motivated to avoid forward-simulating to correctly evaluate backward photos: forward simulation during the assessment of backward photos is detrimental to performance. And indeed, participants were slower in assessing backward photos.

However, still there might be another interpretation of these data. The setup of Experiment 2, in which 2 photos are presented at encoding, one representing the onset and the other the offset of the central part of an action, still might implicitly suggest the need to run a forward mental simulation of the action from onset to offset still frames of the action. Then the simulation might not be spontaneous and mandatory, but activated by the experimental conditions, that implicitly suggest the necessity to run a mental simulation to perform the task. By presenting just one photo

one takes away this implicit nudge and thus if one obtains the same results (faster forward evaluations of congruency) then one can conclude that a forward mental simulation occurs anytime one sees an action also when frozen in time, as in a single still frame. If so, this provides even stronger evidence in favour of the thesis that seeing an action mandatorily triggers a forward kinematic mental simulation of that action.

5. Experiment 3: Implicit evaluation is faster for forward than backward photos (implicit task after observation of a single photo)

The experiment was a replication of Experiment 2 except for the stimulus given to the participants at encoding and some differences in the type of photos presented at test. At encoding, participants observed just one still frame extracted from the middle of the videos used in Experiments 1 and 2, and then, at test, they were invited to decide whether the photo presented depicted a scene congruent or not with the action seen in the photo at encoding. In addition, at test the participants encountered two types of backward photos, depicting the action at two different temporal distances *before* the central part of the action (backward condition), and two different types of forward photos, depicting the action at two different temporal distances *after* the central part of the action (forward condition). In this way, for both conditions there was a photo closest in time to the original action (close backward/forward conditions) and a photo distant in time to the original action (distant backward/forward conditions). The prediction was that evaluation of forward photos should be faster than evaluation of backward photos. No predictions were made about the temporal distances: if assessment is faster for the close forward condition compared to the distant forward condition this would speak in favour of a very fast and short simulative process; otherwise, the forward effect might suggest the existence of a more general mechanism, able to internally generate the representation of the ongoing action for a slightly longer amount of time. Furthermore, the close and the distant photos had different perceptual features (in both backward and forward conditions). In other words, at test close photos were more perceptually similar to the

original photo seen at encoding, while distant photos were less perceptually similar. An effect of temporal distance, then, might also reveal a role of perceptual processes in evaluating the congruency of the photos seen at test with those seen at encoding.

5.1 Method

5.1.1. Participants

Thirty-six students from Turin University (6 males and 30 females, mean age = 21.5, $SD = 4.04$) voluntarily took part in the experiment in exchange of a course credit.

5.1.2. Material

The action videos from which the experimental material was taken were the same as in Experiments 1 and 2. However, 3 of the 10 videos were randomly selected to select the filler stimuli representing actions that were different from those observed at encoding. Four more videos already used in previous studies were also used to select filler stimuli (Iani et al., 2021). Each of the remaining 7 videos (stimuli numbered 1,5,6,7,8,9,10 in Table 1) was cut into three same-length parts. From each video, we extracted the central frame of the middle part to obtain the single photo to present at both encoding and test. From each video we extracted two backward frames and two forward frames to use at test. The “distant backward photo” and the “close backward photo” were the middle frame of the first part and the first frame of the middle part of the whole video, respectively. The “distant forward photo” and the “close forward photo” were the middle frame of the third part and the last frame of the middle part of the whole video, respectively. In Figure 4 are examples of stimuli obtained by the video “Eat a hamburger”.

Figure 4

The central frame extracted from the video “Eat a hamburger” presented to the participants at encoding (Central frame) and the relative photos presented at test (Close Backward photo, Distant Backward photo and Close Forward photo, Distant Forward photo).



Central frame



Close Backward



Close Forward



Distant Backward



Distant Forward

5.1.3. Design and procedure

Design and procedure were the same as in Experiment 2, with the exception that, at test, each scenario was combined with the 4 test photos depicting the action backward or forward in time, both “close” to or “distant” from the scenario represented in the central frame seen at encoding (see the full set of stimuli in the Supplementary Materials). The same scenario was also paired with 4 filler photos depicting different actions from those portrayed in the central photo seen at encoding. We used a within-subjects design; each participant observed the same scenario eight times during

all the experiment (4 times associated to photos requiring YES responses, i.e., close/distant forward and close/distant backward and 4 time associated to photos fillers requiring NO responses). Stimuli were presented in a random order, with the only constraint that the same scenario was not presented in two consecutive trials. The task of participants was to evaluate whether each of these testing photos was congruent with the scenario observed in the previous photo. Nothing in the instructions given to the participants indicated that their task was to infer the goal of the actor in the frame.

Participants were instructed as follows:

Thank you for participating in this experiment on how we understand the unfolding of an action over time. You'll see a photo that represents an action in progress. Soon after each photo a central fixation cross will appear, followed by another photo. Your task is to decide whether the photo depicts a scene congruent or not with the action seen in the initial photo. Answer out loud by saying YES if you think it is congruent, by saying NO if you think it is not.

5.2. Results

The results of 8 of the 36 participants were outliers considering the accuracy in the backward (backward close and backward distant) and in the forward (forward close and forward distant) conditions. The accuracy of eight participants with forward photos fell below the lower limit. For the remaining 28 participants we verified whether the Type of test photo (forward vs backward) and the Type of distance (close vs distant) influenced participants' reaction times (log-transformed) for correct responses using a linear mixed effect regression model. Following the guidelines in the psycholinguistic literature (Barr et al., 2013), we started from the theoretical full model (by including the maximal structure of random effects supported by the design) and then applied to this model the *step function* in the *lmerTest* package (Kuznetsova et al. 2017) to find the final mixed model in which the factors with non-significant effects were removed (see Appendix B, for the theoretic full model and the final model). The final model revealed a significant effect of Type of

photo ($\beta = .065$, $SE = .014$, $t = 4.49$, $p < .001$) on log-transformed reaction times, whereas we did not find either a main effect of Type of distance ($\beta = .012$, $SE = .011$, $t = 1.07$, $p = .29$) or an interaction ($\beta = .017$, $SE = .016$, $t = 1.06$, $p = .29$)². This means that participants were faster in correctly assessing forward photos (a mean of 1283 ms, $SD = 350$) compared to backward photos (a mean of 1562 ms, $SD = 555$), regardless of whether the photos represented something very close to the original action or something more distant in time. Figure 5 shows RTs for correct evaluations in backward and forward conditions. As for Experiments 1 and 2, participants' assessment was more uncertain in backward condition than in forward condition (see the RTs dispersion in the two experimental conditions in Figure 5).

To strengthen what one can infer when interpreting the null effect of temporal distance, we also ran a Bayesian factors analysis (see Rouder et al., 2009; we used JASP software – version 0.14.1.0, and a Cauchy prior with a location parameter of 0 and scale parameter of 0.707). By comparing RTs for close (a mean of 1421.5, $SD = 466.5$) and distant (a mean of 1423, $SD = 411.7$) conditions, testing the H0 (no differences between the two conditions) and the H1 (differences between the two conditions), the result of the Bayesian t -test provides substantial support for the null hypothesis ($BF_{01}=4.98$), being almost 5 times more likely to occur under the null hypothesis, compared with the alternative hypothesis (see Jarosz & Wiley, 2014).

Also, participants' likelihood of error for forward photos and backward photos did significantly differ across backward and forward conditions. Since linear mixed models generated always a singular fit, we ran a simple linear regression model using `lm()` function in R. We detected a significant effect of Type of photo ($\beta = .89$, $SE = .23$, $t = 3.87$, $p < .001$), suggesting that the likelihood of error was greater in assessing congruent backward photos than forward photos. Accuracy was greater for forward photos than for backward photos (a mean of 1, $SD = 0$ and a

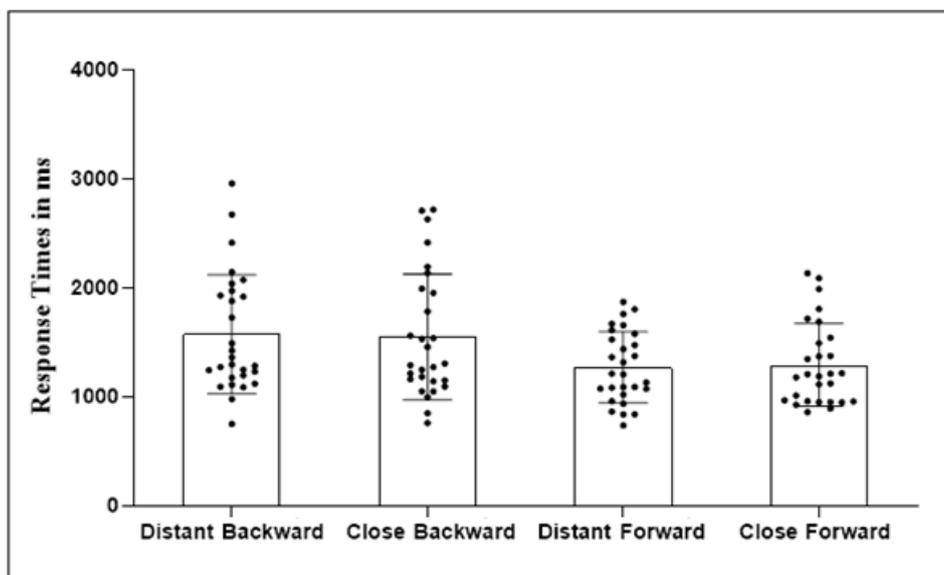
² We run the same analyses also on the total number of participants obtaining the same results (see the file “Experiment 3: Additional statistical analyses” in the Supplementary Materials).

mean of .87, $SD = .16$, respectively). We did not detect either a significant effect of Type of distance ($\beta = .11$, $SE = .23$, $t = .46$, $p = .64$), or an interaction ($\beta = .11$, $SE = .33$, $t = .33$, $p = .74$).

Table A3 in the Appendix summarizes means and standard deviations for response times and errors by the participants in the experiment.

Figure 5

Response times (in ms) and standard deviations for correct evaluations in Experiment 3.



5.2.1. Discussion

The participants in Experiment 3, like those in Experiment 2, were invited to assess the congruency (or not) of new still frames of actions (taken forward and backward in time) with actions seen at encoding but, differently from participants in Experiment 2, at encoding they observed just one single photo, rather than pairs of photos, for each action. If people spontaneously run a forward kinematic mental simulation of the action seen even from a single static picture, we expected the congruency of forward frames to be evaluated faster than that of backward frames. The results confirmed the prediction, thus providing further evidence in favour of the thesis that forward

simulation triggered by photo observation can prime and affect the evaluation of a subsequent action.

6. General Discussion

Participants in our investigation saw pairs of photos depicting the first and the last frame of the middle part of an action video. Then they saw a photo depicting either the middle frame of the first part of the whole video (backward photo) or the middle frame of the third part of the whole video (forward photo). When their task was to decide whether the photo depicted what happened before or after the action seen in the photos, they were faster in evaluating forward photos than backward photos (Experiment 1). This result indicates that people mentally simulate the development of an action in time, even when they see just still frames from the central part of the action. The mental simulation triggered by action observation made it possible to faster identify forward parts (compared to backward parts) of the action that were not presented and not seen. Since in this first experiment participants were explicitly asked to state whether the photos at test represented backward or forward moments of the action, this instruction might have induced an intentional mental simulation of the action. To avoid any explicit suggestion that might induce a mental simulation, participants' task in Experiment 2 was to decide whether the photos at test depicted an action that was congruent with the action seen in the photos at encoding or not. Also in this case participants were still faster in evaluating forward photos compared to backward photos. Crucially, the same effect was detectable also when at encoding participants observed just a single frame rather than pairs of photos representing the onset and the offset of actions (Experiment 3). These results speak in favour of the mandatory nature of forward simulation.

While the facilitation of forward photos had already been detected in one of our previous studies (Iani et al., 2021), the present investigation, by presenting at encoding pairs of photos (the first and the last frame of each action stimulus, Experiment 1-2) or even a single photo (Experiment 3) instead of videos, rules out alternative explanations of the forward effect as merely due to a

different temporal distance of presentation between the initial part of the video and the backward photo and the final part of the video and the forward photo. The results of Experiment 3 also excluded a possible role of perceptual processes in evaluating the congruency of the photos seen at test with those seen at encoding. Photos of the action close to or distant from the moment of the action portrayed in the single photo presented at encoding were perceptually different from each other, and close photos were more like the original photo than distant photos. Nonetheless, participants were faster in correctly assessing the congruency of forward compared to backward photos, regardless the distance in time and the degree in perceptual similarity.

Within our theoretical framework, kinematic mental simulations play a crucial role in action observation: they allow an observer to infer the action that an actor is near to perform. Kinematic simulations represent the mechanism *through which* the observer infers the actor's goal; the observed kinematic features trigger a kinematic mental simulation specific of a certain goal. Consistent with this assumption, the kinematic features of a specific action are predictive of its end and goal (see, Castiello, 2005) because they depend on the actor's intention: different intentions result in different kinematic patterns (e.g., Sartori et al., 2009). Studies revealing that scripts may affect recognition memory suggest a possible alternative account of our results. In one of those studies, the participants watched consecutive slides describing a man in a kitchen and at recognition, when they encountered verbal descriptions of the slides, they misrecognised non-presented but schema-consistent actions, such as "breaking eggs" (Yamada & Itsukushima, 2013, but see also Garcia-Bajos & Migueles, 2003). These results suggest the possibility that the participants in our experiments inferred a goal for the actor from contextual information (McDonough et al., 2020) rather than from kinematic mental simulations. The results of Experiment 1 (forward photos recognized as forward actions and backward photos recognized as backward actions) and the results of Experiments 2 and 3 (faster assessment of forward photos than of backward photos) are not compatible with a script account. The rationale is that the different

(backward and forward) photos from a same action involve the same actor interacting with the same object in the same environment, and they are all equally plausible in that environment.

Globally considered, the results of the present investigation, along with those of our previous study, enforce the assumption that when seen the still frame of an action motor information is mandatorily processed and anticipated through mental simulation, in line with claims that during action observation humans “look forward ... (by an) automatic, implicit, and non-reflexive simulation mechanism” (Gallese, 2005, p.117).

The present investigation represents a further step for an in-depth understanding of the nature of mental simulation from action observation and, although it was beyond its aim to establish whether mandatory motor simulation is involved in cognitive phenomena other than action observation, it may help to differentiate processes involved in the simulation of cognitive phenomena such as language comprehension (see, for a review, Fischer & Zwaan, 2008), memory retrieval (see, for a review, Iani, 2019) and mind reading (see, for a review, Goldman, 2009). Simulative mechanisms can be conceived, at least, either as conscious and explicit imagination of others' action or as subpersonal cognitive mechanisms unfolding without conscious control (Jeannerod & Pacherie, 2004). The evidence provided by our experiments leads to conceive simulation from action observation as a subpersonal process, probably involving activation of mirror neurons or, more generally, resonance systems (e.g., Fadiga et al., 1995; Rizzolatti et al., 1996; Rizzolatti et al., 2000). Put simply, one's motor system reverberates or resonates in one's encounter with others; it is more than likely that “whenever we face situations in which exposure to others' behavior require a response by us, be it active or simply attentive, we seldom engage ourselves in an explicit, deliberate interpretive act. Our understanding of a situation most of the time is immediate, automatic, and almost reflex like” (Gallese, 2005, p. 102).

Gallagher (2007) purports that the concept of “simulation” cannot be legitimately applied to these processes and suggests to consider mental states deriving from action observation as a specific case of “enactive perception” (see also Hurley, 1998; Noë, 2004). “Enactive” means that this

process involves sensory-motor skills, rather than being just sensory input/processing. Perceiving other people's actions is not tantamount to passive reception of information from the world, but rather to an active and embodied process requiring an engagement with the world. According to this interpretation, the mirror activation does not correspond to the initiation of a "simulation". Rather, the mirror activation would be part of a "direct intersubjective perception" (Gallagher, 2007; De Jaegher, 2009), or "relatively smart perception" (Gallagher, 2008), of what other people are doing. However, in Gallagher's view and terminology, motor resonance processes are not a sort of process that would be sufficient to anticipate the forward pattern of movements. Within this debate, the present results concerning the anticipatory nature of the process involved in action observation justify the use of the terms "simulation from action observation". In this view mirror activations, besides resonating what is observed, "simulate" something that is not actually perceived.

Several authors claimed that these subpersonal mechanisms constitute a simulation of others' goal and intentions (Gallese, 2001; Gallese & Goldman, 1998). As a matter of principle, our 'forward mental simulation' interpretation is not necessarily inconsistent with the 'goal anticipation' interpretation (e.g. Gallese, 2001; Gallese & Goldman, 1998). In our theoretical framework anticipation of an action's goal might occur because of the underlying mental simulation triggered by the observed action (e.g., Iani et al., 2018, 2021; Aglioti et al., 2008; Urgesi et al., 2010). Evidence that motor activation in the observers is maximal when participants observe incomplete actions, rather than during the observation of the final state of an action (Urgesi et al., 2010) seems to enforce the assumption that motor representations provide an internal model of the ongoing action (Wilson & Knoblich, 2005), that is a simulative process. Further, evidence suggests that corticospinal excitability is facilitated by combined action observation and explicit mental simulation of a basketball free throw (Wright et al., 2018). However, it remains unclear the point to which forward simulation can represent the upcoming action. From the current study it is possible to deduce that such simulation slightly anticipates the upcoming action (i.e., 250 ms). Does this simulation represent also further steps of the observed action? As pointed out by Catmur (2015),

mirror neurons may be involved in lower-level processes of action perception, but there is no evidence yet to support their involvement in higher-level intention. Future studies need to explore more in depth this issue.

Limitations

A limit of the present investigation is that it was not planned to explain what exactly is responsible for the simulation trigger. Since we used actions on objects as stimuli, we are unable to disentangle what exactly in the image is triggering the cognitive processes underlying the detected effects. However, we hypothesize three possible non-mutually exclusive explanations. Action simulation could be triggered by (1) the actor's kinematics (e.g., Sartori et al., 2009), (2) the object's kinematics/features (Reynaud et al., 2019), (3) the actor's gaze (Ambrosini et al., 2015). Several studies suggest that the kinematic features of a specific action (explanation 1) are predictive of its end and goal (see, Castiello, 2005). Furthermore, studies have shown that the kinematics of a single action depends on the actor's intention: different intentions result in different kinematic patterns (e.g., Sartori et al., 2009). Parallel to this, several studies demonstrated that observers are able to "read" the kinematic information conveyed by the observed actions. For instance, motor information conveyed in action pictures or videos is sufficient for the identification of the specific actor's intention (see, e.g., Manera et al., 2015) as well as the action fate (e.g., if a basket shot success or not, Aglioti et al., 2008). Parallel to this, several studies also suggest a role of objects' knowledge (explanation 2). A meta-analysis of neuroimaging studies suggested that the activation of the action-observation network varied according to whether participants in the studies were invited to observe non-tool-use actions (e.g., a grasping movement) or tool-use-actions (e.g., pouring the water from the bottle; Reynaud et al., 2019); observing actions with tools recruits also non-motor neural resources (i.e., the PF area, within the left inferior parietal lobe). In particular, studies have revealed that actions involving tools use lead individuals first to generate appropriate mechanical actions through technical reasoning, then to select the appropriate motor actions via the motor-control system (Osiurak et al., 2020). Finally, several studies have demonstrated that we also

use cues like eye gaze (explanation 3) to anticipate others' actions (Ambrosini et al., 2015). Given the evidence in favour of each of these three potential explanations, future studies are needed to explore what exactly in the image can trigger the forward action simulation.

Future studies could provide further evidence for the mandatory nature of forward simulations by detecting forward effects similar to those in Experiments 2 and 3 of the present investigation in tasks where the direction of action is even more irrelevant (e.g., a task in which participants are asked to identify whether some details of the test photo, such as the colour of the background, exactly match the stimulus photo). They could also gather interesting information by examining what happens when the stimuli are presented in an order that contradicts the natural flow of the action. In other domains, such as assessing moral responsibility to an actor after reading a short story, the order in which information is presented is important. One study showed that participants found it easier to create a kinematic mental model on which to base judgments of actor responsibility when they read a scenario in which events were presented as they occurred in real life (Limata et al., submitted). Similarly, when participants see two pictures showing the same action at two different times, the reversed order of presentation of the action could affect the ability to simulate the action in progress.

6. Conclusions

The present investigation aimed at shedding light onto the nature of the mental representations triggered by action observation. We hypothesized that action observation elicits a mandatory mental simulation of the action unfolding in time. The participants in the first two experiments observed pairs of photos depicting the same action in different consecutive moments. Then at test they observed a photo depicting either the action backward (backward photo) or forward (forward photo) in time. As the results revealed, participants were faster in deciding that the photo depicted what happened after the action seen in the photos than in deciding that it depicted what happened before (Experiment 1) and they were faster in deciding that a forward photo depicted a scene congruent

with the action seen in the photos than in deciding that a backward photo depicted a congruent scene (Experiment 2). In a final experiment (Experiment 3) we replicated the results of Experiment 2 but showing participants a single photo rather than pairs of photos, thereby suggesting how our brain mentally simulate the implied action from static pictures. While further studies could use concomitant tasks to gain more stringent evidence, taken together, these results suggest that forward simulation stemming from action observation is mandatory.

Acknowledgments

The Supplementary Materials and the data reported in this paper are archived at the following database: <https://osf.io/w6d5y/>

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

References

- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*, 1109-1116. DOI: <https://doi.org/10.1038/nn.2182>
- Ambrosini, E., Pezzulo, G., & Costantini, M. (2015). The eye in hand: Predicting others' behavior by integrating multiple sources of information. *Journal of Neurophysiology*, *113*, 2271-2279. DOI: <https://doi.org/10.1152/jn.00464.2014>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*, 400-404. DOI: <https://doi.org/10.1111/j.1460-9568.2001.01385.x>
- Cardellicchio, P., Sinigaglia, C., & Costantini, M. (2013). Grasping affordances with the other's hand: A TMS study. *Social Cognitive and Affective Neuroscience*, *8*, 455-459. DOI: <https://doi.org/10.1093/scan/nss017>
- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*, 1148-1167. DOI: <https://doi.org/10.1016/j.neuroimage.2009.12.112>
- Castiello, U. (2005). The neuroscience of grasping. *Nature Reviews Neuroscience*, *6*, 726-736. DOI: <https://doi.org/10.1038/nrn1744>
- Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition*, *36*, 426-433. DOI: <https://doi.org/10.1016/j.concog.2015.03.012>

- De Jaegher, H. (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition*, *18*, 535-542. DOI: <https://doi.org/10.1016/j.concog.2008.10.007>
- Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45-60. DOI: <https://doi.org/10.1016/j.cortex.2017.01.016>
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*, 2608-2611. DOI: <https://doi.org/10.1152/jn.1995.73.6.2608>
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*, 825-850. DOI: <https://doi.org/10.1080/17470210701623605>
- Gallagher, S. (2007). Simulation trouble. *Social Neuroscience*, *2*, 353-365. DOI: <https://doi.org/10.1080/17470910601183549>
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*, 535-543. DOI: <https://doi.org/10.1016/j.concog.2008.03.003>
- Gallese, V. (2001). The ‘shared manifold’ hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies*, *8*, 33-50.
- Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*, 23-48. DOI: <https://doi.org/10.1007/s11097-005-4737-z>
- Gallese, V. (2007). Before and below ‘theory of mind’: Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 659-669. DOI: <https://doi.org/10.1098/rstb.2006.2002>

- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Sciences*, 2, 493-501. DOI: [https://doi.org/10.1016/S1364-6613\(98\)01262-5](https://doi.org/10.1016/S1364-6613(98)01262-5)
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609. DOI: <https://doi.org/10.1093/brain/119.2.593>
- Garcia-Bajos, E., & Migueles, M. (2003). False memories for script actions in a mugging account. *European Journal of Cognitive Psychology*, 15, 195-208. DOI: <https://doi.org/10.1080/09541440244000102>
- Goldman, A. I. (2009). Mirroring, simulating and mindreading. *Mind & Language*, 24, 235-252. DOI: <https://doi.org/10.1111/j.1468-0017.2008.01361.x>
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377-442. DOI: <https://doi.org/10.1017/S0140525X04000093>
- Guasti, M. T., Pagliarini, E., & Stucchi, N. (2017). Language, reading, and motor control: Get rhythm!. *Biolinguistics*, 11, 395-406.
- Heyes, C., & Catmur, C. (2020). What happened to mirror neurons?. *Perspectives on Psychological Science*. DOI: <https://doi.org/10.1177/1745691621990638>
- Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.
- Iani, F. (2019). Embodied memories: Reviewing the role of the body in memory processes. *Psychonomic Bulletin & Review*, 26, 1747-1766. DOI: <https://doi.org/10.3758/s13423-019-01674-x>
- Iani, F., Foadelli, A., & Bucciarelli, M. (2019). Mnemonic effects of action simulation from pictures and phrases. *Acta Psychologica*, 194, 37-50. DOI: <https://doi.org/10.1016/j.actpsy.2019.01.012>

- Iani, F., Limata, T., Bucciarelli, M., & Mazzoni, G. (2020). Children's kinematic false memories. *Journal of Cognitive Psychology*, *32*, 479-493. DOI: <https://doi.org/10.1080/20445911.2020.1796686>
- Iani, F., Limata, T., Mazzoni, G., & Bucciarelli, M. (2021). Observer's body posture affects processing of other humans' actions. *Quarterly Journal of Experimental Psychology*, DOI: <https://doi.org/10.1177/17470218211003518>.
- Iani, F., Mazzoni, G., & Bucciarelli, M. (2018). The role of kinematic mental simulation in creating false memories. *Journal of Cognitive Psychology*, *30*, 292-306. DOI: <https://doi.org/10.1080/20445911.2018.1426588>
- Jarosz, A. F. & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2-9. DOI: <https://docs.lib.purdue.edu/jps/vol7/iss1/2>
- Järvilehto, T., Nurkkala, V. M., & Koskela, K. (2009). The role of anticipation in reading. *Pragmatics & Cognition*, *17*, 509-526. DOI: <https://doi.org/10.1075/pc.17.3.02jar>
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind & Language*, *19*, 113-146. DOI: <https://doi.org/10.1111/j.1468-0017.2004.00251.x>
- Kent, C., & Lamberts, K. (2008). The encoding–retrieval relationship: retrieval as mental simulation. *Trends in Cognitive Sciences*, *12*, 92-98. DOI: <https://doi.org/10.1016/j.tics.2007.12.004>
- Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, 1-16. DOI: <https://doi.org/10.3758/s13428-021-01546-0>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1-26. DOI: <http://dx.doi.org/10.18637/jss.v082.i13>
- Limata, T., Iani, F., & Bucciarelli, M. (submitted). Story order in attribution of moral responsibility.

- Mahon, B. Z., & Caramazza, A. (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cognitive Neuropsychology*, *22*, 480-494. DOI: <https://doi.org/10.1080/02643290442000446>
- Manera, V., Ianì, F., Bourgeois, J., Haman, M., Okruszek, Ł. P., Rivera, S. M., Robert, P., Schilbach, L., Sievers, E., Verfaillie, K., Vogeley, K., von der Lue, T., Willems, S., & Becchio, C. (2015). The multilingual CID-5: A new tool to study the perception of communicative interactions in different languages. *Frontiers in Psychology*, *6*, 1724. DOI: <https://doi.org/10.3389/fpsyg.2015.01724>
- McDonough, K. L., Costantini, M., Hudson, M., Ward, E., & Bach, P. (2020). Affordance matching predictively shapes the perceptual representation of others' ongoing actions. *Journal of Experimental Psychology: Human Perception and Performance*, *46*, 847-859. DOI: <https://doi.org/10.1037/xhp0000745>
- Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Science*, *364*, 1273-1280. DOI: <https://doi.org/10.1098/rstb.2008.0314>
- Noë, A. (2004). *Action in perception*. MIT press.
- Osiurak, F., Badets, A., Rossetti, Y., Lesourd, M., Navarro, J., & Reynaud, E. (2020). Disembodying (tool-use) action understanding. *Neuroscience and Biobehavioral Reviews*, *114*, 229-231. DOI: <https://doi.org/10.1016/j.neubiorev.2020.03.020>
- Pozzo, T., Papaxanthis, C., Petit, J. L., Schweighofer, N., & Stucchi, N. (2006). Kinematic features of movement tunes perception and action coupling. *Behavioural Brain Research*, *169*, 75-82. DOI: <https://doi.org/10.1016/j.bbr.2005.12.005>
- Quesque, F., & Coello, Y. (2015). Perceiving what you intend to do from what you do: evidence for embodiment in social interactions. *Socioaffective Neuroscience & Psychology*, *5*, 28602. DOI: <https://doi.org/10.3402/snp.v5.28602>

- R Development Core Team (2021). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna.
- Reynaud, E., Navarro, J., Lesourd, M., & Osiurak, F. (2019). To watch is to work: A review of NeuroImaging data on tool use observation network. *Neuropsychology Review*, *29*, 484-497. DOI: <https://doi.org/10.1007/s11065-019-09418-3>
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-141. DOI: [https://doi.org/10.1016/0926-6410\(95\)00038-0](https://doi.org/10.1016/0926-6410(95)00038-0)
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. *The New Cognitive Neurosciences*, *2*, 539-552.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225-237. DOI: <https://doi.org/10.3758/PBR.16.2.225>
- Sartori, L., Becchio, C., Bara, B. G., & Castiello, U. (2009). Does the intention to communicate affect action kinematics? *Consciousness and cognition*, *18*, 766-772. DOI: <https://doi.org/10.1016/j.concog.2009.06.004>
- Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, *11*, 349-355. DOI: <https://doi.org/10.1016/j.tics.2007.06.005>
- Urgesi, C., Maieron, M., Avenanti, A., Tidoni, E., Fabbro, F., & Aglioti, S. M. (2010). Simulating the future of actions in the human corticospinal system. *Cerebral Cortex*, *20*, 2511-2521. DOI: <https://doi.org/10.1093/cercor/bhp292>
- Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*, 217-232. DOI: <https://doi.org/10.1080/13506280143000403>

- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9, 625-636.
<https://doi.org/10.3758/BF03196322>
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131, 460-473. DOI: <https://doi.org/10.1037/0033-2909.131.3.460>
- Wright, D. J., Wood, G., Eaves, D. L., Bruton, A. M., Frank, C., & Franklin, Z. C. (2018). Corticospinal excitability is facilitated by combined action observation and motor imagery of a basketball free throw. *Psychology of Sport and Exercise*, 39, 114-121. DOI: <https://doi.org/10.1016/j.psychsport.2018.08.006>
- Yamada, R., & Itsukushima, Y. (2013). The effects of schema on recognition memories and subjective experiences for actions and objects. *Japanese Psychological Research*, 55, 20-32.
DOI: <https://doi.org/10.1111/jpr.12016>

Appendix A.

Table A1

Mean and standard deviation of response times (in ms) and errors as a function of type of photo in Experiment 1

	Type of photo	
	Backward	Forward
RTs	2001 <i>SD= 825</i>	1630 <i>SD= 539</i>
Errors	.19 <i>SD= .22</i>	.06 <i>SD= .09</i>

Table A2

Mean and standard deviation of response times (in ms) and errors as a function of type of photo in Experiment 2

	Type of photo	
	Backward	Forward
RTs	1405 <i>SD= 559</i>	1193 <i>SD= 411</i>
Errors	.05 <i>SD= .13</i>	0 <i>SD= 0</i>

Table A3

Mean and standard deviation of response times (in ms) and errors as a function of type of photo and type of distance factors in Experiment 3

	Type of photo ~ Type of distance			
	Distant Backward	Close Backward	Distant Forward	Close Forward
RTs	1575 <i>SD= 545</i>	1550 <i>SD= 576</i>	1272 <i>SD= 325</i>	1293 <i>SD= 380</i>
Errors	.13 <i>SD= .14</i>	.13 <i>SD= .17</i>	0 <i>SD= 0</i>	0 <i>SD= 0</i>

Appendix B

The theoretical full model for each experiment and the final model found by the step function in the package lmerTest.

Experiment 1:

1) *Full model*: Reaction Times (Log) ~ Type of photo + (1 + Type of photo|Subjects) + (1 + Type of photo|Items)

2) *Final model*: Reaction Times (Log) ~ Type of photo + (1 + Type of photo|Subjects)

Experiment 2:

1) *Full model*: Reaction Times (Log) ~ Type of photo + (1 + Type of photo|Subjects) + (1 + Type of photo|Items)

2) *Final model*: Reaction Times (Log) ~ Type of photo + (1|Subjects)

Experiment 3:

1) *Full model*: Reaction Times (Log) ~ Type of photo*Type of distance + (1 + Type of photo*Type of distance|Subjects) + (1 + Type of photo*Type of distance|Items)

2) *Final model*: ReactionTimes_log ~ Type_Photo + Type_Distance + (Type_Photo | Subject) + Type_Photo:Type_Distance