Cognition 153 (2016) 33–42

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Article

Action-outcome learning and prediction shape the window of simultaneity of audiovisual outcomes

Andrea Desantis *, Patrick Haggard

Institute of Cognitive Neuroscience, University College London, London, UK

ARTICLE INFO

Article history:
Received 26 August 2015
Revised 4 March 2016
Accepted 8 March 2016

Keywords:
Agency
Voluntary action
Action-outcome learning
Action-outcome prediction
Time perception
Audiovisual binding

ABSTRACT

To form a coherent representation of the objects around us, the brain must group the different sensory features composing these objects. Here, we investigated whether actions contribute in this grouping process. In particular, we assessed whether action-outcome learning and prediction contribute to audiovisual temporal binding. Participants were presented with two audiovisual pairs: one pair was triggered by a left action, and the other by a right action. In a later test phase, the audio and visual components of these pairs were presented at different onset times. Participants judged whether they were simultaneous or not. To assess the role of action-outcome prediction on audiovisual simultaneity, each action triggered either the same audiovisual pair as in the learning phase ('predicted' pair), or the pair that had previously been associated with the other action ('unpredicted' pair). We found the time window within which auditory and visual events appeared simultaneous increased for predicted compared to unpredicted pairs. However, no change in audiovisual simultaneity was observed when audiovisual pairs followed visual cues, rather than voluntary actions. This suggests that only action-outcome learning promotes temporal grouping of audio and visual effects. In a second experiment we observed that changes in audiovisual simultaneity do not only depend on our ability to predict *what* outcomes our actions generate, but also on learning the delay between the action and the multisensory outcome. When participants learned that the delay between action and audiovisual pair was variable, the window of audiovisual simultaneity for predicted pairs increased, relative to a fixed action-outcome pair delay. This suggests that participants learn action-based predictions of audiovisual outcome, and adapt their temporal perception of outcome events based on such predictions.

1. Introduction

Our environment comprises complex objects characterized by auditory, visual, and other sensory features that are processed by partially independent brain areas. The brain must be able to appropriately group information deriving from different senses in order to identify these objects and generate a unified perceptual experience of our surroundings.

Multisensory grouping partly depends on the co-occurrence in time of different sensations (Meredith, Nemitz, & Stein, 1987): multisensory interactions are stronger when two or more modalities are perceived as occurring simultaneously (Alais, Newell, & Mamassian, 2010; Stein & Meredith, 1990). Previous research showed that multisensory temporal simultaneity is important for guiding our actions by enabling fast and accurate responses (Colonius and Diederich, 2004; Colonius and Arndt, 2001; Frens, Van Opstal, & van der Willigen, 1995; Stein & Meredith, 1990).

A more radical view of grouping reverses the causal relation between perception and action, suggesting that action drives the perceptual processes that produce multisensory grouping (cf. Lewkowicz & Ghazanfar, 2009; James, 1890; Petrini, Russell, & Pollick, 2009; Piaget, 1963). Although, the specific role of action in multisensory processing remains under-researched, there is evidence that actions may shape such processes. Past research showed that action processes strongly mediate the perception unimodal stimuli. Indeed, sensory events caused and predicted by one's own actions are attenuated compared to stimuli that are externally-generated and predicted by sensory cues (cf, sensory attenuation, Blakemore, Wolpert, & Frith, 2000; Cardoso-Leite, Mamassian, Schütz-Bosbach, & Waszak, 2010; Hughes, Desantis, & Waszak, 2013). Research on sensory attenuation has also shown that this effect depends largely on action processes involved in the

* Corresponding author.
E-mail address: aerdna.desantis@gmail.com (A. Desantis).

http://dx.doi.org/10.1016/j.cognition.2016.03.009
0010-0277/© 2016 The Authors. Published by Elsevier B.V.

preparation of actions and in the neural prediction of the specific sensory consequences that our actions produce[1] (Stenner, Bauer, Heinze, Haggard, & Dolan, 2014; Blakemore et al., 2000; Cardoso-Leite et al., 2010; Wolpert, 1997). Taken together, these studies suggest that action processes mediate the transformation of physical stimulation into perceptual experience (Wilson & Knoblich, 2005). In line with these notion, the present study investigated whether the processes involved in action planning and action-outcome prediction also shape perceptual multimodal grouping. Supporting evidence for this hypothesis comes from a recent study showing that active exploration of audiovisual objects enhances memory and subsequent recognition of these objects compared to passive observation (Butler, James, & James, 2011).

Previous studies investigating the influence of action on perceptual experience, generally focussed on action outcomes confined to a single sensory modality. In real life, however, most actions produce multisensory effects. For example, speaking produces auditory, kinaesthetic and tactile inputs. Consequently, when preparing/executing an action the motor system might predict several sensory outcomes (i.e., auditory, tactile sensations) to occur *together* as a *common* outcome of our motor command. Importantly, with the term prediction we refer to the prediction of the content/identity of an action-outcome. In other words, we hypothesized that when we predict that our actions generate a specific combination of a sound and a visual input, we might tend to group these inputs into a simultaneous multisensory percept. In line with this notion, the *unity assumption* of multisensory perception (Jackson, 1953; Vatakis & Spence, 2007; Welch, 1999) states that sensory events that "go together" are experienced as simultaneous, even if they are slightly asynchronous (Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence, 2007). Here, we test the hypothesis that action-outcome learning and prediction (i.e., the prediction of the content/identity of an action-outcome) facilitates the process of audiovisual binding. That is, action-outcome learning and prediction may enlarge a hypothetical "temporal window" within which the different multisensory components of an action outcome are perceived to be simultaneous. In this paper, we will refer to this concept as the Window of Audiovisual Simultaneity (WAS). This grouping process may be crucial, suggesting that the action system contributes to the unity and coherence of our perceptual experience (i.e., active exploration would help us create a unified and coherent representation of the external world, e.g., Piaget, 1963). Secondly, it might be essential to develop a healthy sense of agency. Indeed, when we generate outcomes composed of different features, the brain must be able to selectively bind the sensory components of these outcomes and not others, to prevent erroneous self-attributions.

To assess the role of action-outcome learning and prediction on multisensory binding, we conducted three experiments (one of them is reported in supplementary material) using a *mismatch paradigm,* in which the match/mismatch between predicted and actual action outcomes was varied (for similar methods see Baess, Widmann, Roye, Schröger, & Jacobsen, 2009; Cardoso-Leite et al., 2010; Desantis, Mamassian, Lisi, & Waszak, 2014). Participants were presented with two audiovisual pairs. They learned that one pair followed a left hand action, and the other pair followed a right hand action. Audio and visual inputs were presented simultaneously, but the interval between the action and the audiovisual pair was jittered. In a later test phase, each action could trigger either the same audiovisual pair as in the initial learning phase (the 'predicted' pair), or the pair associated with the other action ('unpredicted' pair). The latter case created a mismatch between predicted and actual action outcome. Importantly, the association of audio and visual components within each pair remained unbroken throughout the experiment: match/mismatch occurred between action and outcome, and never within the components of the outcome itself. In the test phase, the interval between the audio and visual components of each pair varied, and participants judged whether they were presented simultaneously or not. We hypothesised that learning a specific action-outcome relation would temporally bind the audio and visual components within the outcome pair (Fig. 1).

As a consequence of this process, the audio and visual components of the predicted outcome should more readily be experienced as simultaneous, even when slightly asynchronous, compared to the unpredicted outcome. To clarify whether this multisensory perceptual binding was indeed driven by *action*-outcome prediction, we compared a condition in which the participants voluntarily triggered the outcome through their own action, and a condition where the participants made no actions, but the outcomes were predicted by visual cues, with the same latency and probability relations as the action condition. We expected to observe no change of the WAS between predicted and unpredicted pairs when these were associated to visual cues and not actions.

In Experiment 2, we investigated whether the relation between action-outcome learning and temporal binding within the audiovisual outcome might itself be temporally tuned. That is, when participants learn the relation between an action and a multisensory outcome, they may also learn the time window within which the audio and visual components of the outcome should be bound together. Specifically, we hypothesised that a reliable temporal delay between an action and a predicted outcome should lead to a narrower temporal window for binding the predicted components of the outcome, relative to a variable delay. Evidence in support of this hypothesis comes from studies on sensory attenuation. Notably, research demonstrated that sensory attenuation of predicted action-outcome (by predicted action-outcome we mean the prediction of *what* outcome an action generates) occurs specifically around the time at which participants' expect the predicted outcome to occur (Bays, Wolpert, & Flanagan, 2005). Moreover, previous studies showed that prior experiences can recalibrate the window of audiovisual grouping (see Fujisaki, Shimojo, Kashino, & Nishida, 2004; Roseboom & Arnold, 2011; Spence & Squire, 2003; Vroomen, Keetels, de Gelder, & Bertelson, 2004), but it remains unclear whether action-outcome learning can induce such changes. Strategic tuning of the WAS could play an important role in parsing sensory events into those that are self-caused, and those that are externally-generated. For example, incorrectly setting too wide an action-based window for multisensory binding might lead to erroneous self-attribution of multimodal events, in a manner reminiscent of delusions of control.

2. Experiment 1

2.1. Materials and methods

2.1.1. Participants

Sixteen volunteers (12 women, average age = 21.28 years, *SD* = 3.78 years) were tested for an allowance of £ 7.5/h or course credit. Participants completed the experiment in two sessions on separate days (see supplementary material for inclusion criteria). All had normal or corrected-to-normal vision and hearing and were naïve as to the hypothesis under investigation. They all gave written informed consent. The experiments were conducted with ethical committee permission.

[1] In the present manuscript with the term *action-outcome prediction* or *prediction* we refer to the ability to predict *what* outcome an action generates. In other words, we refer to the prediction of the content or identity of an action-outcome. We use, instead, the term *temporal expectation* or *expectation* to refer to the ability to anticipate the *time* onset of action-outcomes.