Cognition 153 (2016) 19–32

Contents lists available at ScienceDirect

Cognition

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

Observing others stay or switch – How social prediction errors are integrated into reward reversal learning

Niklas Ihssen [a,*], Thomas Mussweiler [b], David E.J. Linden [c]

[a] Department of Psychology, Durham University, Durham DH1 3LE, UK
[b] Department of Psychology, University of Cologne, 50931 Cologne, Germany
[c] School of Psychology & MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, CF10 3AT, UK

ARTICLE INFO

Article history:
Received 15 August 2014
Revised 12 February 2016
Accepted 17 April 2016

Keywords:
Reversal learning
Social influence
Reward
Prediction error
Similarity

ABSTRACT

Reward properties of stimuli can undergo sudden changes, and the detection of these 'reversals' is often made difficult by the probabilistic nature of rewards/punishments. Here we tested whether and how humans use social information (someone else's choices) to overcome uncertainty during reversal learning. We show a substantial social influence during reversal learning, which was modulated by the type of observed behavior. Participants frequently followed observed conservative choices (no switches after punishment) made by the (fictitious) other player but ignored impulsive choices (switches), even though the experiment was set up so that both types of response behavior would be similarly beneficial/detrimental (Study 1). Computational modeling showed that participants integrated the observed choices as a 'social prediction error' instead of ignoring or blindly following the other player. Modeling also confirmed higher learning rates for 'conservative' versus 'impulsive' social prediction errors. Importantly, this 'conservative bias' was boosted by interpersonal similarity, which in conjunction with the lack of effects observed in a non-social control experiment (Study 2) confirmed its social nature. A third study suggested that relative weighting of observed impulsive responses increased with increased volatility (frequency of reversals). Finally, simulations showed that in the present paradigm integrating social and reward information was not necessarily more adaptive to maximize earnings than learning from reward alone. Moreover, integrating social information increased accuracy only when conservative and impulsive choices were weighted similarly during learning. These findings suggest that to guide decisions in choice contexts that involve reward reversals humans utilize social cues conforming with their preconceptions more strongly than cues conflicting with them, especially when the other is similar.

1. Introduction

Adaptive behavior depends on learning and retaining associations between specific stimuli or responses on the one hand and positive or negative outcomes (reward or punishment) on the other. In a complex and dynamic environment organisms must also adequately respond to sudden changes in those associations and re-learn established contingencies. A widely used experimental tool to study this process in animals and humans is reversal learning (Cools, Clark, Owen, & Robbins, 2002; Dias, Robbins, & Roberts, 1996; Jones & Mishkin, 1972). In a typical setup, human participants learn to choose one of two simple visual stimuli by receiving monetary rewards for correct responses (stimulus A) and being punished by monetary loss for incorrect responses (stimulus B). After a variable number of trials, these contingencies are reversed so that the participant will be rewarded for choosing B and be punished for choosing A. Trial-by-trial choices in this task can be predicted by the algorithms of reinforcement learning models which are based on the calculation of reward prediction errors (Jocham, Neumann, Klein, Danielmeier, & Ullsperger, 2009).

1.1. Social information and decision-making

Critically, in real-life situations learning of reward contingencies is not only achieved by trial-and-error and reward prediction errors but also by social learning, that is, by observing the choices of other agents who are exposed to the same or similar decisional contexts. In the majority of everyday choice situations (e.g. choosing between alternative products or services) social information is readily available either through behavioral observation of others or

* Corresponding author at: Department of Psychology, Durham University, Wolfson Building, Queen's Campus, Stockton-on-Tees, TS17 6BH, UK.
E-mail address: niklas.ihssen@durham.ac.uk (N. Ihssen).

http://dx.doi.org/10.1016/j.cognition.2016.04.012
0010-0277/

through active gathering of information (e.g. consumer reviews). Observational factors can be expected to become especially important if well-established behavioral choice routines need to be revised because the expected outcome is not received or experienced as less rewarding. In such situations the possibility of a change in the underlying reward probabilities (e.g., the quality of the usually preferred product/service has changed) will evoke decisional uncertainty which is a potent trigger for 'social reality testing', that is, the reliance on others to resolve ambiguity (Festinger, 1950). The literature to date has ignored whether information about others' choices affects responding to sudden changes in reward properties of a stimulus as implemented in the reversal learning task. This is surprising given the evidence that other basic cognitive processes, such as perceptual judgments, can profoundly be shaped by social influence (Asch, 1956; Baron, Vandello, & Brunsman, 1996).

Social influence can be governed, on the one hand, by socio-normative mechanisms, originating from the influenced person's motivation to gain social approval if the influencing person is present (as in Asch's classic line discrimination studies). On the other hand, it can also arise in the absence the influencing person and social pressure, being motivated by informational needs (Deutsch & Gerard, 1955) and the wish to resolve ambiguity to optimize one's outcomes. Such informational social influence is likely to operate in choice decisions involving uncertain rewards and a few studies have begun to document social influences on probabilistic reward learning. However, these studies used fixed (Biele, Rieskamp, Krugel, & Heekeren, 2011; Burke, Tobler, Baddeley, & Schultz, 2010) or gradually changing (Behrens, Hunt, Woolrich, & Rushworth, 2008; Cooper, Dunne, Furey, & O'Doherty, 2012) reward contingencies rather than a setup involving unexpected reversals.

1.2. Predictions for the use of social information during reversal learning

The primary goal of the present studies was to explore the use of social information (observed choices by another agent) during reversal learning, specifically, how such social influence is mediated by the (i) type of observed choice behavior (conservative versus impulsive) and (ii) the similarity of the observed agent.

The differentiation between conservative and impulsive choices during reversal learning arises as a result of the task-inherent combination of probabilistic reward and possibility of reversals. In other words, even if reward contingencies have not changed, correct choices are occasionally punished by monetary losses (so-called Probabilistic Errors, ProbErrs). Consequently, after each punishment occurring against the background of correct responses an individual has to decide whether to switch to the other stimulus (taking the punishment as indicator of reversed contingencies) or whether to stay with their previous choice (taking the punishment as a ProbErr). Accordingly, choices in trials that immediately follow ProbErrs and reversals can be classified as reflecting either a *conservative* or an *impulsive* type of choice behavior. Stay responses correspond to conservative choices as the agent relies on accumulated information about a specific choice option – which has been gathered across a number of trials before the unexpected punishment – rather than trying a new option. This can be seen in analogy to an *exploitative* decision-making strategy in multi-armed bandit problems, in which multiple choice options with varying pay-offs are available (Cohen, McClure, & Yu, 2007). Conversely, switch responses can be seen as *impulsive* choices (Fineberg et al., 2010), reflecting an abrupt change of choice behavior based on single events without consideration of the previous choice history. Importantly, adaptive behavior during reversal learning requires the dynamic use of both types of behavior. Although impulsive responses manifest as errors in the trial(s) after ProbErrs (=post-ProbErr choice) they lead to correct choices in the trials(s) after true reversal events. Vice versa, conservative responses increase accuracy after ProbErrs but lead to incorrect choices ('persevera tions') after reversal events. The key question addressed in the present framework was whether *observing* someone else making conservative choices affects our own choices differently than observing someone else making impulsive choices.

Diverging effects for observed conservative versus impulsive choices can be predicted from findings about the biased use of information during individual decision-making. Thus, it is possible that learners take into account only social information conforming to their preconception or expectation about the correct versus incorrect stimulus (established before the other player's choice is observed). This preconception is based on the learner's more frequently chosen stimulus in a given reversal episode and thus usually corresponds to a bias towards conservative choices made by the observed other. Such selective use of social information would parallel a 'confirmation' bias described in the context of individual decision-making (Nickerson, 1998). Conversely, a social influence bias towards the other's impulsive choices may arise if observational reversal learning is expectation-free but driven by the higher saliency of impulsive (switch) responses occurring against the stream of standard (non-switch) choices between two reversals.

Apart from the type of observed choice behavior, the present studies aimed to examine social influence on reversal learning as a function of perceived similarity of the observed agent. Similarity has been shown to influence different cognitive processes across a wide range of phenomena, including decision-making (Kahneman & Miller, 1986). Similarity is also effective in modulating a variety of social behaviors, ranging from the experience of vicarious reward (Mobbs et al., 2009) to cooperative behavior (Mussweiler & Ockenfels, 2013). Pertinent to the present work, the behavior and opinions of similar versus dissimilar others are more likely to be imitated (Guéguen & Martin, 2009). Moreover, requests from similar others are more likely to be complied with than requests from dissimilar others, suggesting that similarity directly affects the degree of social influence (Burger, Messian, Patel, del Prado, & Anderson, 2004).

With regard to the role of similarity, we thus hypothesized that any bias in the following of behavioral patterns of the observed agent should be exaggerated (i.e. social learning rates should be increased) if this person shares a characteristic feature with the observing agent.

2. General methods

2.1. The social reversal learning task

On each trial of the present task, participants observed the response of a (fictitious) other player before they were required to make their own choice. Reversal learning performance was assessed in two different blocks, examining choice behavior without (private/baseline block, Fig. 1A) and with (social block, Fig. 1B) exposure to the choices made by a (fictitious) previous participant. In both blocks, participants learned to choose one of two simultaneously presented colors[1] ('blue' and 'green') by receiving monetary reward (+1 pence [p]) or punishment (−1p) contingent on their choice (e.g. +1p for 'blue' and −1p for 'green').

After a variable number of trials unknown to the participants, reward/punishment contingencies were reversed so that the previously rewarded stimulus was now punished and vice versa (−1p

[1] For interpretation of color in Fig. 1, the reader is referred to the web version of this article.