Cognition 119 (2011) 394–402

Contents lists available at ScienceDirect

Cognition

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

Optimistic biases in observational learning of value

A. Nicolle*, M. Symmonds, R.J. Dolan

Wellcome Trust Centre for Neuroimaging at UCL, 12 Queen Square, London WC1N 3BG, UK

ARTICLE INFO

Article history:
Received 10 May 2010
Revised 1 February 2011
Accepted 2 February 2011
Available online 26 February 2011

Keywords:
Learning
Social cognition
Decision making
Value
Reward

ABSTRACT

Action-outcome contingencies can be learnt either by active trial-and-error, or vicariously, by observing the outcomes of actions performed by others. The extant literature is ambiguous as to which of these modes of learning is more effective, as controlled comparisons of operant and observational learning are rare. Here, we contrasted human operant and observational value learning, assessing implicit and explicit measures of learning from positive and negative reinforcement. Compared to direct operant learning, we show observational learning is associated with an optimistic over-valuation of low-value options, a pattern apparent both in participants' choice preferences and their explicit post-hoc estimates of value. Learning of higher value options showed no such bias. We suggest that such a bias can be explained as a tendency for optimistic underestimation of the chance of experiencing negative events, an optimism repressed when information is gathered through direct operant learning.

1. Introduction

Many instances of everyday learning rely upon trial-and-error. Here, a decision-maker samples between alternative actions and risks unfavorable outcomes in the early stages of learning, when action-outcome contingencies are unknown. Learning can also occur through observing the successes and failures of others, enabling us to acquire knowledge vicariously. Indeed, the benefits of observational learning are ubiquitous in nature. For example, a hungry animal can avoid the energy costs incurred in active sampling of optimal feeding locations by observing actions and outcomes of conspecifics. A proliferation of customer review websites epitomizes the utility of learning through the positive and negative experiences of others so obviating our own need for expensive decisions. In this way, observational learning is recognized as supporting "locally adaptive behaviors without incurring the costs associated with individual learning" (Boyd & Richerson, 1988, p. 30).

Surprisingly, the efficacy of observational learning has been rarely studied in the context of human value learning. Empirical evidence in animals attests to the fact that rewarded behavior is promoted, and punished behavior diminished, in passive observers (e.g. Bandura, 1971; Dawson & Foss, 1965; Heyes & Dawson, 1990; Mineka & Cook, 1988; Weigl & Hanson, 1980). For example, budgerigars show imitation of rewarded behaviors but a diminution of such behavior if the observed consequences are not salient, suggesting that vicariously conditioned responses are goal-directed and not a mere mimicry of an observed action (Heyes, 1994; Heyes & Saggerson, 2002). However, despite these data, evidence for the effectiveness of observational learning is inconsistent. Church (1959) found that rats observing lights predicting a shock to a model do not generalize these contingencies to their own risk preferences.

Several critical differences can be highlighted between vicarious and active value learning, which may lead to differences in information acquisition. One factor is motivation, of key importance in Bandura's (1977) social learning theory, given that passive observers do not directly incur costs or benefits during learning. Our emotional responses, enhanced when we act and experience

---

* Corresponding author.
E-mail address: a.nicolle@fil.ion.ucl.ac.uk (A. Nicolle).

0010-0277 © 2011 Elsevier B.V. Open access under CC BY license.
doi:10.1016/j.cognition.2011.02.004

outcomes ourselves, motivate our learning and decision-making (e.g. Schwarz & Clore, 1983). Anticipated emotions may also increase attention and an incentive to learn, and are likely to be greatest when actively learning. Alternatively, our emotions can potentially distract from, or “crowd out”, our goals (Loewenstein, 1996), or bias our memory for the frequency of past events (cf. emotional biases of eyewitness testimonies, e.g. Loftus, 1996), both of which could disrupt learning. Consistent with this “dark side of emotion”, individuals with decreased emotional responses for outcomes of risky decisions can show more advantageous decision-making (Shiv, Loewenstein, & Bechara, 2005).

Operant and observational learning differ in how attention is directed during learning. An actor’s ability to selectively sample an environment facilitates learning of an existing ‘region of uncertainty’ (Cohn, Atlas, & Ladner, 1994). Observers, on the other hand, lack this sampling control, making learning potentially inefficient. Observational learning may require a more explicit, declarative acquisition of knowledge, which may not be necessary given the procedural nature of operant learning (Howard, Mutter, & Howard, 1992; Kelly, Burton, Riedel, & Lynch, 2003; Willlingham, 1999). Neumann (1990) has argued that perception needs to be concurrently tied with action to influence subsequent behavior. fMRI studies, and assessments of learning deficits in Parkinson’s patients, support a functional dissociation of declarative or observational learning from non-declarative, procedural learning (Ostlund & Balleine, 2007; Poldrack et al., 2001; Shohamy et al., 2004). Furthermore, while explicit knowledge acquisition may be subject to distraction by other motivations, implicit learning of action-outcome associations may be less vulnerable to distraction (Neumann, 1990). From these considerations it is reasonable to predict superior learning through action than through observation.

In this study, our aim was to make a controlled comparison between active and observational learning in the context of human probabilistic value learning. Thus, we implemented a learning task where individuals learnt either by active sampling (with associated reward and punishment) or by passive observation. We assessed learning efficacy as shown by goal-directed choices and individuals’ explicit estimates of value. All aspects of the tasks, save for the critical factor of self versus other choice, were matched across two modes of value learning. Specifically, differences in attention and information were controlled, as participants could track the same sequences of outcomes in both learning conditions, as was motivation to learn, since participants earned money according to learning performance in both conditions.

2. Experiment 1

2.1. Participants

In this first experiment we recruited 17 healthy participants, screened for neurological or psychological disorders. Participants failing to reach a criterion of 60% accuracy by the end of each session, when choosing between the 80/20 probability of winning pair, were excluded from further analysis, given a performance level barely exceeding chance (i.e. 50% accuracy) and was considered as a failure to engage sufficiently with the task. This was the case only for one participant, leaving 16 participants for the full analysis (nine female, mean age 23.8 yrs, SD 3.0). Participants provided informed consent, according to UCL Research Ethics Committee approved procedures.

2.2. Procedure

Participants completed two sessions on consecutive days. In the first (the ‘actor session’), participants made choices between four stimuli (letters from Agathodaimon font), presented in different pairs on each trial, while concurrently attempting to learn the probability of winning from each. Participants were made aware that each stimulus was associated with a discrete and constant probability of winning (p{win}), and outcomes of each stimulus were drawn independently on every trial. Outcomes of chosen and unchosen stimuli were then shown sequentially, with yellow and red boxes indicating winning and losing outcomes, respectively. Critically, these outcomes directly influenced participant’s earnings for the actor session (with £1 awarded for each chosen win from 10 randomly selected trials). Participants were instructed to choose the stimulus with the highest p{win} on every trial in order to maximize earnings.

On day 2 (the ‘observer session’), participants learned the values of a novel set of stimuli (stimulus sets were balanced between sessions and across participants). We gave an instruction that this time participants would observe choices made previously by another participant, along with their associated outcomes. Participants were not provided with any information about this other participant, but were informed that these were real choices made by a different individual in a prior session. Participants were informed that, although they could learn from the outcomes of observed choices, these outcomes would *not* influence their own earnings for the observer session. Unknown to them, participants observed the sequence of choices they had made in their previous actor session, although now with visually novel stimuli. The two sessions were, therefore, matched in terms of the information from which they learned. Observer sessions were completed on day 2 in order to reduce memory for previous choice sequence. To match for motor responses, observers indicated the observed choice on each trial with a button-press. Since learning could not be measured in these observation trials, because a free choice is not made, we introduced test trials to assess learning in both actors and observers. These comprised nine blocks of trials (test blocks) at regular intervals throughout learning. Here, free choices were made by both actors and observers in the absence of outcome feedback (to prevent further learning).

Fig. 1 illustrates exemplar learning and test trials and indicates the sole difference between actors and observers at the time of choice. Participants played a total of 324 trials per session (i.e. actor or observer). There were 12 trials in each of nine learning blocks, allowing for six