Cognition 133 (2014) 128–139

Contents lists available at ScienceDirect

Cognition

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

Demonstrator skill modulates observational aversive learning

Ida Selbing *, Björn Lindström, Andreas Olsson

Karolinska Institute, Division of Psychology, Nobels väg 9, 171 65 Solna, Sweden
Stockholm Brain Institute, Retzius väg 8, 171 65 Solna, Sweden

ARTICLE INFO

Article history:
Received 9 October 2013
Revised 10 June 2014
Accepted 13 June 2014

Keywords:
Observational learning
Avoidance
Skill
Reinforcement learning

ABSTRACT

Learning to avoid danger by observing others can be relatively safe, because it does not incur the potential costs of individual trial and error. However, information gained through social observation might be less reliable than information gained through individual experiences, underscoring the need to apply observational learning critically. In order for observational learning to be adaptive it should be modulated by the skill of the observed person, the demonstrator. To address this issue, we used a probabilistic two-choice task where participants learned to minimize the number of electric shocks through individual learning and by observing a demonstrator performing the same task. By manipulating the demonstrator's skill we varied how useful the observable information was; the demonstrator either learned the task quickly or did not learn it at all (random choices). To investigate the modulatory effect in detail, the task was performed under three conditions of available observable information; no observable information, observation of choices only, and observation of both the choices and their consequences. As predicted, our results showed that observable information can improve performance compared to individual learning, both when the demonstrator is skilled and unskilled; observation of consequences improved performance for both groups while observation of choices only improved performance for the group observing the skilled demonstrator. Reinforcement learning modeling showed that demonstrator skill modulated observational learning from the demonstrator's choices, but not their consequences, by increasing the degree of imitation over time for the group that observed a fast learner. Our results show that humans can adaptively modulate observational learning in response to the usefulness of observable information.

1. Introduction

Learning to avoid potentially dangerous events through trial and error can be a painful experience. Learning the same thing indirectly by observing the actions of others can be both safe and efficient. For example, watching a pair of boxers go head to head in a fight is a less painful way of learning how to avoid the harmful consequences of boxing compared to learning by stepping into the ring oneself. One way to minimize harm during boxing is to learn which side is best to lean in response to a straight punch. This can be learned in a safe manner through observation of someone else, a demonstrator, by attending to (1) which side the other person, the demonstrator, chooses to lean and (2) the consequences of that choice. Learning a response or choice observationally will here be referred to as observational learning or, when speaking of aversive learning, observational avoidance learning. Learning an association, rather than a response, through observation will be referred to as observational associative learning.

* Corresponding author at: Nobels väg 9, 171 65 Solna, Sweden. Tel.: +46 762501865.
E-mail addresses: ida.selbing@ki.se (I. Selbing), bjorn.lindstrom@ki.se (B. Lindström), andreas.olsson@ki.se (A. Olsson).

http://dx.doi.org/10.1016/j.cognition.2014.06.010
0010-0277/

Observational learning of choices and actions is likely to be modulated by the demonstrator's degree of skill, defined here as his or her ability to minimize negative or maximize positive consequences. The primary reason for this is that the demonstrator's skill is predictive of the degree to which his/her choices reflect the underlying contingency between choice and consequence, affecting the usefulness of observing the choices. In many real life learning situations where we can observe others, information is often limited. We might not know which choices were available to the demonstrator and consequences of the choices might be delayed. Also, we often lack knowledge of the skill or experience of those we observe, for example, when the demonstrator is unknown to us. It is therefore of great importance to be able to evaluate and use various sources of observable information critically. Recent research has shown a modulatory effect of demonstrator skill on observational learning in humans (Apesteguia, Huck, & Oechssler, 2007) where participants' choices were influenced more when demonstrators appeared more skilled than themselves, and observational learning in other (non-human) social animals (Kendal, Rendell, Pike, & Laland, 2009) where food patch choices in fish were influenced more by successful than unsuccessful conspecifics. This research has, however, not connected with research describing the processes underlying observational associative learning, such as associative learning of fear in humans (Olsson & Phelps, 2007) as well as social avoidance learning in other animals (Kavaliers, Choleris, & Colwell, 2001). Furthermore, it is unknown how the skill of the demonstrator modulates observational learning from choices and consequences, respectively.

Several studies have shown that observational learning and other forms of social learning can outperform individual learning (Feldman, Aoki, & Kumm, 1996; Kameda & Nakanishi, 2003; Merlo & Schotter, 2003; Rendell et al., 2010). In particular, social learning is theorized to be especially advantageous when negative consequences can be costly (Dewar, 2004; Kendal, 2004; Webster & Laland, 2008), such as in dangerous environments (Coolen, van Bergen, Day, & Laland, 2003; Galef, 2009). Despite this, studies of observational learning have almost invariably focused on learning within the appetitive domain (e.g. Apesteguia et al., 2007; McElreath et al., 2008; Merlo & Schotter, 2003). In contrast, avoidance learning depends on reinforcers in the form of punishing aversive events (Dayan & Balleine, 2002) or the rewarding omittance of an aversive event (Rescorla, 1969). For instance, the boxer in our previous example learns the correct defense when choices are punished by a hit, a naturally aversive consequence, a primary reinforcer. Previous research of observational associative learning in humans have commonly used primary reinforcers, such as shocks (Olsson, Nearing, & Phelps, 2007), whereas studies of observational learning have used secondary reinforcers, such as money (e.g. Burke, Tobler, Baddeley, & Schultz, 2010; Merlo & Schotter, 2003; Nicolle, Symmonds, & Dolan, 2011; Suzuki et al., 2012). Our first aim with the present study was thus to extend the literature on observational learning into the aversive domain using primary reinforcers.

Although observational learning often is thought of as safe and efficient, observational (social) information can be outdated or inaccurate and observational learning should thus be applied critically (Enquist, Eriksson, & Ghirlanda, 2007; Kendal, Coolen, van Bergen, & Laland, 2005). For instance, available theories on social learning strategies (Laland, 2004; Schlag, 1999) suggest that copying should be more common if the demonstrator is successful. In humans, this theory is supported by empirical research where explicit feedback of the participant's and the demonstrator's overall performances is given (Apesteguia et al., 2007; Mesoudi, 2008; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012) and by studies showing that people take more advice from an experienced or trained advisor than a novice (Biele, Rieskamp, & Gonzalez, 2009; Sniezek, Schrah, & Dalal, 2004). To our knowledge, no human studies have investigated if and how demonstrator skill modulates observational avoidance learning when explicit information of skill level is not given. Our second aim was thus to study the impact of the demonstrator's skill on observational learning in a paradigm where skill level could only be inferred by observation of the demonstrator's choices and consequences.

To target the aims of our study, we adopted an experimental paradigm previously used with secondary reinforcers that allowed us to disentangle observational learning from choices and consequences (Burke et al., 2010). In our design (described in detail in Fig. 2), participants learned a sequential probabilistic two-choice task using naturally aversive reinforcers. In addition to making their own choices, participants were also at times able to observe a demonstrator that learned the same task. The choices were reinforced by electric shocks so that one out of a pair of choices was punished more often than the other. To investigate the influence of available information on observational learning trials belonged to one of three observational learning conditions with varying amounts of available observable information: (1) no observable information, individual learning (No Observation), (2) observable information of the demonstrator's choices (Choice Observation) and (3) observable information of both the demonstrator's choices and the consequences of

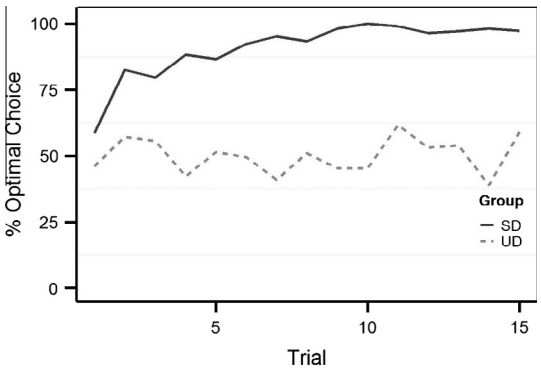

Fig. 1. Mean demonstrator performance (defined as the percentage of optimal choices) per trial in each block differed between groups; the performance level increased rapidly for the SD group while the performance level remained at chance for the UD group.

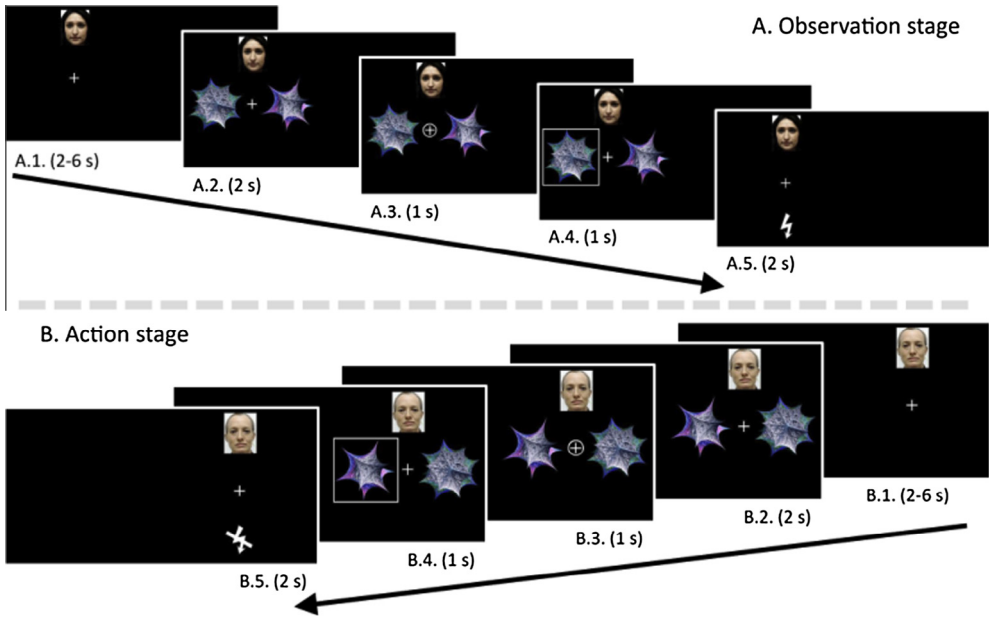

Fig. 2. A.1., B.1.: Varying intertrial interval (ITI) displaying a fixation-cross (2–6 s). A.2., B.2.: Choice stimuli presentation (2 s). A.3.: Circled fixation-cross used as a "go-signal" requiring participants to press the up-arrow in order to later observe the demonstrator's choices or both choices and consequences (ensuring that participants were attentive to the display, 1 s). B.3.: Circled fixation-cross used as a "go-signal" requiring participants to make a choice with the left or right arrow (1 s). A.4.: Demonstrator's choice was indicated with a white frame during Choice Observation and Choice-Consequence Observation (1 s). During No Observation both stimuli were framed. B.4.: Participant's choice indicated with a white frame (1 s). A.5.: A symbol was shown to indicate the consequence of the demonstrator's choice, a shock symbol or crossed shock symbol (Choice-Consequence Observation only) or a simple line when no information of consequences was given (2 s). B.5.: A symbol was shown to indicate the consequence of the participant's choice, a shock symbol or crossed shock symbol. The onset of the shock symbol display coincided with the administration of an electric shock to the participant's wrist (100 ms). If the participant did not make a required response during A.3., both choice stimuli were framed during A.4., and during A.5. (regardless of condition) a non-informative line was shown as a symbol. If the participant did not make a required response during B.3., both choice stimuli were framed during B.4., and during B.5. the consequence was randomized.

those choices (Choice-Consequence Observation). To investigate the effect of demonstrator skill, participants were divided into two groups that either observed a skilled demonstrator (SD) that easily learned the task and had a high level of performance or an unskilled demonstrator (UD) that did not learn but instead selected choices randomly over the entire course of the experiment.

To enhance understanding of how demonstrator skill might modulate observational avoidance learning, we analyzed performance using reinforcement learning (RL) modeling (Sutton & Barto, 1998). This approach allowed us to formalize how a participant's expectations (e.g. of a certain consequence following a specific choice) were updated over time using prediction errors, defined as deviations from expectations (Sutton & Barto, 1981). Prediction errors are believed to guide learning when predictions are violated and have been linked to specific neural markers during learning from both rewarding (Rushworth, Mars, & Summerfield, 2009) and punishing events (Delgado, Li, Schiller, & Phelps, 2008). In addition, the RL-framework has been linked to prediction learning outside of the rewarding/punishing domain, including social learning such as prediction of other people's actions (Burke et al., 2010) and trustworthiness (Behrens, Hunt, Woolrich, & Rushworth, 2008).

We hypothesized that access to observable information would improve performance compared to when no observable information was available (individual learning) similar to the results previously shown by Burke et al. (2010). Moreover, we predicted that participants observing a skilled demonstrator would let observable information guide selection of choices to a higher degree than those observing an unskilled demonstrator. This was expected to be most apparent under the Choice Observation condition where observed choices provide the only observational information regarding the underlying choice-consequence contingency. Since a demonstrator's choices will reflect this contingency only if the demonstrator has an ability to learn, observing choices will be helpful only for the SD group. We used RL-modeling to formalize and describe the separate influences of demonstrator skill on learning from observation of choices and observation of their consequences. Moreover, RL-modeling allowed us to compare different models of observational learning to understand how observable information was used.

2. Materials and methods

2.1. Participants

42 self-reportedly healthy participants were recruited and paid for participation in the experiment approved by the local ethics committee. 2 Participants were excluded

due to performance levels below random leaving a total of 40 participants that were randomly assigned to either the SD (skilled demonstrator) group or the UD (unskilled demonstrator) group (SD: $n = 20$, 13 women, mean age = 23.85 years [SD = 5.76]; UD: $n = 20$, 9 women, mean age = 25.00 years [SD = 5.37]). Participants were informed that they would perform the experiment together with another person and upon arrival they met but did not interact further with a sex-matched confederate (male: age 30, woman: age 26) presented to them as the other participant (i.e. the demonstrator). Before starting, all participants signed an informed consent form and a facial photo was taken for use in the computerized choice task.

2.2. Material

The experiment was presented using E-prime (Psychology Software Tools, Inc., www.pstnet.com). Mild electric shocks consisting of 100 ms DC-pulses (STM200; Biopac Systems Inc.) applied to the left wrist served as the primary reinforcer. The strengths of the electric shocks were individually set to be unpleasant but not painful. Participants used their right hand to press the keyboard keys. For the SD group the choices of the demonstrator were decided using a simple RL model which learned the task relatively quickly (see Appendix A.1.1). For the UD group the demonstrator's choices were random. For a comparison of demonstrator performance between groups, see Fig. 1. Thirty pictures of randomly generated fractals on a black background were used as choice stimuli (180×180 pixels). For each participant 24 of these were randomly picked and randomized into 12 stimulus pairs in which one stimulus was randomly selected as the optimal choice. Of these pairs, 3 were used for a practice block and the other 9 pairs were divided between the remaining 3 blocks. The photo of each participant was cropped and resized to 100×125 pixels before experiment onset. Photos of the demonstrators had previously been taken, cropped and resized in a similar manner.

2.3. Procedure

Participants performed a probabilistic two-choice task adopted from Burke et al. (2010), differing mainly in the number of trials and blocks. Also, whereas the original task included blocks where choices were either punished or rewarded, using monetary feedback, we only used punishment to reinforce choices. Each trial in the setup consisted of an initial *observation stage* during which the demonstrator made his/her choice followed by an *action stage* during which the participant made his/her choice. Each pair of choice stimuli belonged to one of three Observational Learning conditions depending on the amount of available observable information: (1) individual learning (No Observation), (2) observable information of the demonstrator's choices (Choice Observation), (3) observable information of both the demonstrator's choices and the consequences of those choices (Choice-Consequence Observation). Apart from an initial training block, trials were divided into three blocks; each consisting of three pairs of choice stimuli (one for each observational learning condition) displayed 15 times each, resulting in a total of 135 trials. During each stage a photo above the displayed stimuli indicated whose turn it was to make a choice. The demonstrator's choices were shown on the left side of the screen and the participant's choices to the right. For each pair of choices, one was assigned to be the optimal choice and was thus associated with a lower probability of being paired with a shock then the other choice (probabilities were 0.8/0.2 respectively). Fig. 2 displays a detailed description of the stage phases. Each stage started with a fixation cross (duration 2–6 s) followed by presentation of the choice stimuli (duration 2 s). The fixation-cross was then circled (duration 1 s) as a "go-signal". During the observation stage the "go-signal" indicated that participants were required to press a button in order to observe the demonstrator's choices and/or consequences, ensuring that they were attentive to the display. During the action stage the "go-signal" indicated that participants were required to choose one of the stimuli. When the "go-signal" was followed by a required response the demonstrator's or participant's choice (depending on stage) was indicated by framing the chosen stimuli (duration 1 s). If a required response was missing, and during No Observation for the observation stage, both stimuli were framed. Next followed the display of a symbol indicating the consequence (shock or no shock, duration 2 s). For the observation stage during No Observation and Choice Observation the shock and no shock symbols were replaced with a non-informative symbol. For the action stage, the onset of the shock symbol coincided with the administration of an electric shock (duration 100 ms) to the participant's wrist. If the participant did not make a choice during the action stage the risk of a shock was 0.5. Participants were informed that they and the demonstrator were given the same task: to minimize the number of shocks by trying to select the optimal choice in each pair. Participants were also told that the demonstrator could never observe their choices or consequences. Importantly, the participants received no information regarding the skill of the demonstrator, neither before nor during the experiment.

2.4. Reinforcement learning modeling

In order to investigate how the experimental manipulation affected decision making on a trial-by-trial basis we analyzed participant's choices using RL-modeling based on the Q-learning algorithm (Sutton & Barto, 1998; Watkins, 1989) which was extended to include observational learning in a manner similar to that employed by Burke et al. (2010). According to the standard *Q*-learning algorithm, all available choices are associated with action values, *Q*-values, representing their expected consequences. Action values are used as input to the softmax activation function which is used to calculate the probability of making a certain choice. The softmax function assigns the highest probability to the choice with the best expected consequence although this is modulated by a parameter, β, which controls the function's tendency to exploit or explore data. During individual learning, the action value at trial t associated with the choice being made is updated proportional to the difference between

the expected and actual consequence, the prediction error $\delta_{consequence}$, and a learning rate, $\alpha_{individual}$:

$$Q_{choice}(t+1) = Q_{choice}(t) + \alpha_{individual} * \delta_{consequence|choice} \quad (1)$$

After a repeated number of trials, action values can be said to reflect running estimates of the consequences of the choices being made. Observational learning was modeled by including two observational prediction errors: $\delta_{obs.\ conseq.}$, the difference between the expected and obtained *consequence* following the demonstrator's choice, and $\delta_{obs.\ choice}$, the difference between the expected and observed *choice* of the demonstrator. At each trial (where applicable, see Section 2.3.), observational learning precedes individual learning. The prediction error following observation of the consequences of the demonstrator's choice affects the action values similarly as during individual learning using an observational learning rate, $\alpha_{obs.\ conseq.}$:

$$Q_{obs.\ choice}(t+0.5) = Q_{obs.\ choice}(t) + \alpha_{obs.\ conseq.} * \delta_{obs.\ conseq.} \quad (2)$$

These updated action values are then used to compute the probabilities of making each choice. The prediction error following observation of choice is combined with a learning rate parameter, the imitation rate, $\alpha_{imitation}$, to increase the probability of making the same choice as the demonstrator:

$$\rho_{obs.\ choice}(t) = \rho_{obs.\ choice}(t) + \alpha_{imitation} * \delta_{obs.\ choice} \quad (3)$$

The three different learning rates and the inverse temperature parameter β were included as free parameters. For all formulated models free parameters where fitted for each subject over all trials. Models were compared using Akaike Information Criterion weights (AIC weights). AIC is an estimate of the quality of a model that includes a penalty for the number of free parameters the model contains to balance the trade-off between model complexity and goodness-of-fit (Busemeyer & Wang, 2000). AIC weights are used to compare a set of models by calculating the weight of evidence in favor of each model using the models' AIC values (Lewandowsky & Farrell, 2010). Modeling was conducted using R (R Development Core Team, 2012). For details on models and parameter fitting, see Appendix.

3. Results

3.1. Statistical results

Behavioral analyses were carried out on trials where the "go-signal" was followed by a required response (see Section 2.3. or Fig. 2, phases A.3, B.3.) resulting in a mean number of trials per participant of 120.77 (SD = 12.10). Choices were categorized on the basis of performance (optimal/suboptimal), where a choice was optimal if it was the choice associated with the lowest risk of shock, and imitation (imitative/non-imitative), where a choice was imitative if it was the same as the demonstrator's choice (note that imitation is only possible during choices in the observational conditions). For such binary data, logistic regression is the preferred analysis method (Jaeger, 2008) and thus choice data was analyzed using Logistic Generalized Mixed Models (Baayen, Davidson, & Bates, 2008) with by-subject random intercept and full random fixed effects structure (Barr, 2013). All follow up contrasts were adjusted for multiple comparisons using a single-step method based on the multivariate normal distribution (Genz & Bretz, 1999). We began by analyzing effects on performance. We saw no significant difference in overall performance between groups ($p = 0.62$) but we did find a main effect of Observational Learning Condition ($\chi^2(2) = 21.03$, $p < 0.001$). Follow up contrasts showed that performance during No Observation was marginally worse than during Choice Observation ($\beta = -0.18$, $SE = 0.08$, $z = 2.19$, $p = 0.07$) and worse than during Choice-Consequence Observation ($\beta = -0.40$, $SE = 0.09$, $z = 4.59$, $p < 0.001$). Performance during Choice-Consequence Observation was better than during Choice Observation ($\beta = 0.21$, $SE = 0.09$, $z = 2.42$, $p = 0.04$). Our hypothesis that the effect of demonstrator skill would be most apparent under Choice Observation was confirmed by a Group × Observational Learning Condition interaction ($\chi^2(2) = 8.00$, $p = 0.02$) driven mainly by a marginal difference in performance at the Choice Observation condition (follow up contrasts: $\beta = 0.50$, $SE = 0.24$, $z = 2.11$, $p = 0.10$) where the SD group performed better than the UD group, see Fig. 3a. There were no group differences in performance during No Observation or Choice-Consequence

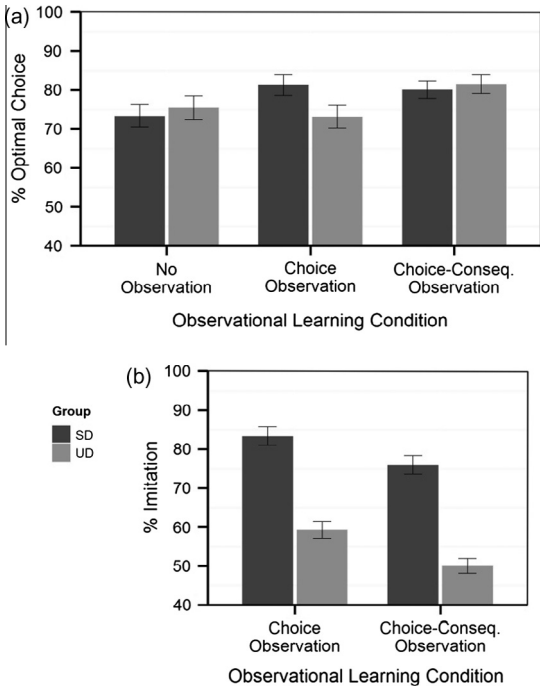

Fig. 3. (a) For both groups, performance levels were higher during Choice-Consequence Observation compared to No Observation. A difference in performance between groups was seen during Choice Observation where the performance level of the SD group was higher compared to the UD group. (b) The SD group displayed an increased level of imitation compared to the UD group over both observational conditions. Also, imitation was higher during Choice Observation compared to Choice-Consequence Observation for both groups.

Observation ($p = 0.93$). Within-group comparisons showed that for the SD group, performance during No Observation was significantly worse than during Choice Observation ($\beta = -0.51$, $SE = 0.19$, $z = -2.63$, $p = 0.02$) and marginally worse than during Choice-Consequence Observation ($\beta = -0.40$, $SE = 0.19$, $z = -2.16$, $p = 0.08$). Performances during Choice Observation and Choice-Consequence Observation did not differ. For the UD group, performance during No Observation did not differ from performance during Choice Observation but was marginally lower than during Choice-Consequence Observation ($\beta = -0.39$, $SE = 0.19$, $z = -2.07$, $p = 0.10$). Performance during Choice Observation was lower than during Choice-Consequence Observation ($\beta = -0.51$, $SE = 0.17$, $z = -2.99$, $p = 0.01$). Thus, our analyses show that both groups performed better when the amount of observable information increased. Observing an unskilled, as compared to a skilled, demonstrator led to impaired performance only during Choice Observation although it was still on par with the performance during No Observation. This suggests that participants were able to proficiently modulate behavior so that observable information was used in a manner that improved performance.

To examine if this modulation of behavior occurred over time, we analyzed the interaction over blocks and found a significant Group × Observational Learning Condition × Block interaction effect ($\chi^2(2) = 7.51$, $p = 0.02$), see Fig. 4a. The interaction was disentangled by analyzing the Observational Learning Condition × Block interaction for each group separately. For the SD group we found a significant interaction effect ($\chi^2(2) = 14.34$, $p < 0.001$) which resulted from a significant increase in performance over blocks during Choice Observation ($\beta = 0.67$, $SE = 0.13$, $z = 5.06$, $p < 0.001$). No such increase was seen for either No Observation or Choice-Consequence Observation. For the UD group, there were no such time dependent effects; neither the Observational Learning Condition x Block interaction ($p = 0.26$) nor the simple effect of Block ($p = 0.10$) were significant. In sum, this shows that participants that observed a skilled, but not an unskilled, demonstrator modulated behavior such that performance improved over time while observing the demonstrators choices but not the consequences.

Next, we analyzed the corresponding effects on imitation. We found a main effect of Group on Imitation ($\chi^2(1) = 101.69$, $p < 0.001$) which was the result of a higher degree of imitation for the SD group compared to the UD group ($\beta = 1.24$, $SE = 0.12$, $z = 10.08$, $p < 0.001$). We also noted a main effect of Observational Learning Condition ($\chi^2(1) = 28.63$, $p < 0.001$) caused by a higher degree of imitation during Choice Observation compared to Choice-Consequence Observation ($\beta = 0.42$, $SE = 0.08$, $z = -5.35$, $p < 0.001$). We found no significant Group × Observational Learning Condition interaction ($p = 0.46$), see Fig. 3b, although we did find a significant Group × Observational Learning Condition × Block interaction ($\chi^2(1) = 4.39$, $p = 0.04$), see Fig. 4b. To further investigate this interaction we analyzed it separately for the two groups. For the SD group, we found a significant Observational Learning Condition × Block interaction effect ($\chi^2(1) = 5.48$, $p = 0.02$). This was due to an increase in imitation over blocks during Choice Observation ($\beta = 0.57$, $SE = 0.14$, $z = 3.94$, $p < 0.001$) while there was no change over blocks for Choice-Consequence Observation. For the UD group, there was

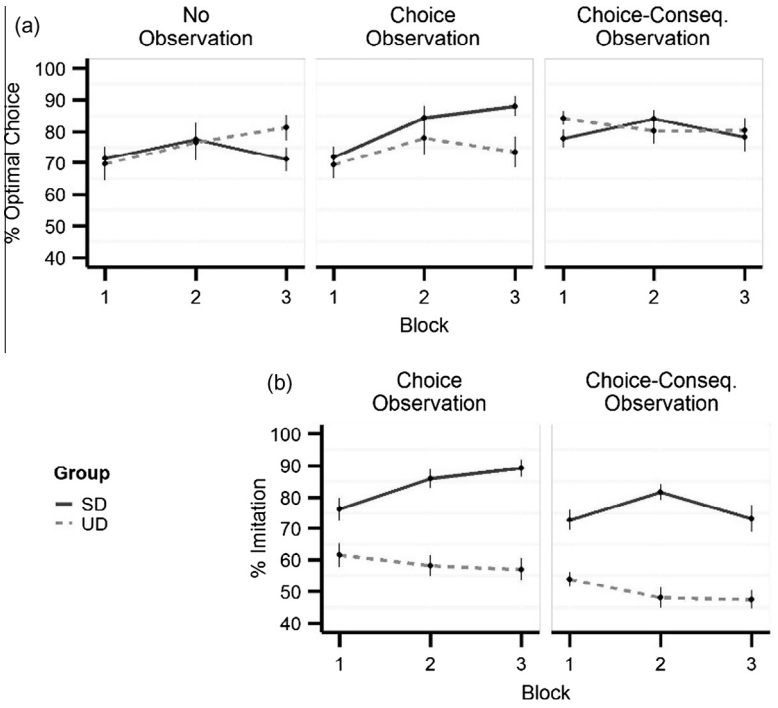

Fig. 4. (a) The only change in performance over blocks was seen during Choice Observation where the SD group increased performance over blocks. (b) For imitation, there was a change over blocks seen in the SD group during Choice Observation where imitation increased over blocks.

no Observational Learning Condition × Block interaction ($p = 0.90$) and no main effect of Block ($p = 0.11$). Thus, parallel to the increase in performance, participants observing a skilled demonstrator over time increased their degree of imitation while observing the demonstrator's choices but not the consequences of those choices. Importantly though, based on the definition of imitation used here, we cannot conclude if this is due to copying of the demonstrators' choices or the fact that two agents that learn the same task are more likely to make the same choices even without copying compared to when one of the agents behaves randomly.

3.2. Model based results

To further disentangle the effects of demonstrator skill on the participants' use of various sources of information we formulated four different RL-models that combined individual and observational learning. The RL-models were based on a basic model of individual learning but included observational learning from choices and consequences. Observational learning from the two observational sources of information was modeled by adding two separate observational learning rates and calculating two separate observational prediction errors. The observational models differed in which sources of information were included and how they were combined. Two of the models included only one source of observable information; CH included observational learning from choices and CO included observational learning from consequences. The remaining two models combined both sources of observable information but did so in different manners during Choice-Consequence Observation when both sources were available; CHCO.H was a hybrid model that combined both sources during Choice-Consequence Observation, CHCO.S kept the sources separated and disregarded observable information of choices during Choice-Consequence Observation (equivalent to the observational RL-model used by Burke et al., 2010). All observational models and a baseline model that included only individual learning were compared by calculating the AIC weight of each model to provide a measure of the evidence for each model given the set of models compared, see Table 1.

Based on the mean AIC weight over participants in each group, CHCO.S was the preferred model for both groups although the mean rank order and number of wins for CO indicates that the model which only included observational learning from observation of consequences also provided a high goodness-of-fit. The preferred model, CHCO.S, was able to predict approximately 68% of participants' choices (69% of group SD and 67% of group UD). Between group comparisons of all the individually fitted free parameters of the preferred model CHCO.S showed a significant difference only for the fitted imitation rate values (two-sample t-test: $t(27.34) = 2.24$, $p = 0.03$) where the SD group had a higher imitation rate than the UD group.

Next, to investigate how observational learning changed over time we formulated two additional models based on the preferred model CHCO.S. These included changes over blocks in either imitation rate, $\alpha_{imitation}$, or observational learning rate, $\alpha_{obs.\ conseq.}$, which was implemented by allowing the respective learning rates to change linearly over the blocks of the experiment (see Appendix for details). Neither of the models yielded any significant change in goodness of fit compared to CHCO.S for any of the two groups. Analyses of variance over the changeable learning rate showed a significant Block × Group interaction for the imitation rate only ($F(1) = 15.14$, $p < 0.001$) not the observational learning rate, see Fig. 5. This interaction was brought about by a significant increase of the imitation rate from first to last block for the SD group (paired t-test: $t(19) = 3.81$, $p < 0.01$, mean increase = 0.20). For the UD group it remained at a relatively low level although separated from zero (one sample t-test: $t(19) = 24.80$, $p < 0.001$) indicating that imitation affected participants' choices in the UD group also at the last block.

To summarize, we found that the model of observational learning that best described participant's behavior in both groups included observational learning from both choices and consequences, but disregarded observable information of choices when consequences were observable. Differences in the fitted model parameters showed that the SD group imitated the demonstrator to a greater extent than the UD group. These differences were the result of an increase of imitation over blocks in the SD group rather than a decrease of imitation in the UD group. These results fit well with our behavioral analyses.

4. Discussion

Learning to avoid dangerous situations is crucial for all animals. If the aversive consequences are sufficiently costly individual learning through trial and error can be very risky making observational learning particularly important in dangerous situations (Webster & Laland, 2008). Here we show for the first time that observational

Table 1
Model comparisons for both groups using mean (M) and standard deviation (SD) of AIC weights over participants and mean rank order (Rank) plus number of wins (Wins) when comparing AIC weights for the included models over each participant. Numbers in bold indicate preferred model for each group according to weight, rank and number of wins separately.

Model	Skilled demonstrator				Unskilled demonstrator			
	M	SD	Rank	Wins	M	SD	Rank	Wins
Individual	0.06	0.13	4.10	2	0.04	0.11	4.30	1
CH	0.05	0.11	3.90	1	0.03	0.05	4.25	0
CO	0.25	0.19	**2.20**	7	0.30	0.19	**1.95**	**8**
CHCO.H	0.19	0.14	2.55	2	0.24	0.18	2.40	4
CHCO.S	**0.45**	**0.34**	2.25	**8**	**0.39**	**0.28**	2.10	7

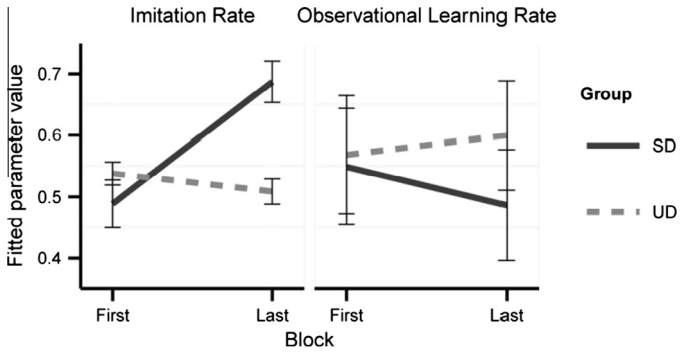

Fig. 5. Temporal changes of parameters associated with observational learning. Individual parameter values are fitted to the first and last block (block 1 and 3) and change linearly over the middle block (block 2). Error bars show standard error of the mean over each group. Note that the imitation rate and the observational learning rate have different, although parallel, functions and comparing their exact values is not meaningful.

avoidance learning is modulated by the skill of the demonstrator affecting performance during observation of choice but not during observation of both choice and consequences. Our results indicate that participants were able to proficiently use observable information. The current study extends previous studies, which have argued that learning from secondary reinforcers is similar to learning from primary reinforcers, both rewarding (O'Doherty, 2004) and punishing (Delgado, Labouliere, & Phelps, 2006) by indicating that this is the case also for observational avoidance learning.

In support of both empirical (Kameda & Nakanishi, 2003; Merlo & Schotter, 2003; Mesoudi, 2008; Morgan et al., 2012) and theoretical (McElreath, Fasolo, & Wallin, 2013; Rendell et al., 2010) research, we showed that observational learning improved performance as compared to individual learning. We found a positive relationship between the amount of available observable information and the participants' ability to make the optimal choice in order to minimize the number of shocks. Importantly, we showed that observable information improved performance irrespective of the level of skill of the demonstrator. Thus, even the group that watched the unskilled demonstrator improved performance when the amount of observable information increased, although only apparent when observation of consequences was available. Performance while observing the demonstrator's choices only was improved when those choices were informative (skilled demonstrator) and was not worsened when choices were uninformative (unskilled demonstrator, random choices). Performance while observing both the demonstrator's choices and consequences was improved regardless of the skill of the demonstrator. This indicates that the participants were able to make use of observable information in an adaptive manner. Importantly, participants did this without explicit information of the skill of the demonstrator.

We used RL-modeling to disentangle the different sources of information involved in observational learning: individual information, observable information of choices and of consequences. Since the participants' task was to learn to select the optimal choice, they needed information about the contingencies between choices and consequences, which could be acquired individually or through observation of the consequences following the demonstrator's choices. Also, provided that the demonstrator performs better than random, information about the contingency between choice and consequences can be acquired through observation of choices. Varying the demonstrator's skill thus affected the observable information so that observation of choices was informative for the SD group, but not the UD group. It is also worth pointing out that the choices of the demonstrator also affect the information gained from observation of the consequences of those choices. For example, if the demonstrator selects one choice more often than the other, which was the case for the SD group, sampling of observed consequences will be biased leading to a difference in the ability to learn the contingencies of the two choices compared to random sampling of observed consequences (Denrell & Le Mens, 2013). Consequently, the SD and UD group in the present study differed not only in how informative the demonstrators' choices were, but also in observational sampling of the consequences of those choices, suggesting that demonstrator skill might modulate learning from observation of both choices and consequences. Our results showed that participants adjusted to the quality of information by modulating learning from observation of choices. However, we saw no modulation of learning from observation of consequences. RL model comparisons indicate that the reason for this could be that when participants observed the consequences of the demonstrator's choices, they only used the observationally learned choice-consequence contingency to guide selection of choices and did not imitate the skilled, or the unskilled, demonstrator. This shows the importance of observational learning of choice-consequence contingencies, a form of learning that appears to influence behavior in humans as early as 12 months of age (Elsner & Aschersleben, 2003).

Furthermore, using temporally changing RL-modeling, we were able to estimate how the modulation of observational learning developed over time. Both groups had an initial low level of imitation and the temporal changes consisted of an increase in imitation over time for the SD group rather than a decrease for the UD group. This effect

was seen even though participants were not offered any explicit information of demonstrator skill or overall performance which has been common in previous studies (e.g. Apesteguia et al., 2007).

It is not clear from our data, however, how the available information was used to guide this modulation of observational learning. We made an attempt using RL-modeling to investigate whether participants modulated observational learning either by learning the value of imitation through directly experienced consequences or the value of the demonstrator's behavior by observing the consequences the demonstrator suffered, but found no support for either of the two possible explanations (see Supplementary Information). An interesting aspect of our results from the temporally changing RL-models is that although we do not see any significant differences in the performance levels for the UD group when comparing individual learning (No Observation) and learning from observation of choices (Choice Observation), the imitation rate does not decrease over time and is separated from zero even at the last block. This indicates that some participants in the UD group might imitate the demonstrator's random behavior even after observing the demonstrator's choices and consequences for several blocks. This inclination to imitate random behavior deserves further investigation. Is it possible that this could stem from a difficulty to learn how to use observable information, either blind imitation (McGregor, Saggerson, Pearce, & Heyes, 2006), stimulus enhancement, where a stimulus is somehow rendered attractive or salient simply by observing a demonstrator interact with it (Heyes, Ray, Mitchell, & Nokes, 2000), or a combination of these explanations? A possible route for a more detailed understanding of how learning is modulated could be to investigate the effects of demonstrator skill on neural activity. For example, by combining RL-modeling and fMRI it would be possible to investigate how demonstrator skill affects the previously demonstrated choice prediction error signal in the dorsolateral prefrontal cortex (Burke et al., 2010).

The results of the present study have important implications for the understanding of observational learning situations in real life where we often lack knowledge of the skill of those we observe and where observable information is not always complete, such as when consequences are delayed. Also, observation of avoidance might be seen as a special case of lack of observable information, because successful avoidance is defined by the omission of the negative consequence. Consequently, it is important not only to be able to use various sources of information, but also to judge their value and use them critically. To return to our initial example, when observing two boxers go head to head in a fight we would predict that it is possible to learn by observing both the winner and the loser by adjusting how the observable information influences behavior. A good idea would be to copy the behavior of the winner but learn from the consequences of both boxers' choices.

Acknowledgements

We thank Armita Golkar for input during the design of the experiment and Tanaz Molapour for assisting us during the data collection. Finally, we thank Cristopher Burke for providing us with information concerning the details of the experimental paradigm upon which our study is based. This research was supported by an Independent Starting Grant (284366; Emotional Learning in Social Interaction project) from the European Research Council to A.O.

Appendix A

Here we describe the details of the RL-modeling and parameter fitting.

A.1. Computational modeling

All models are based on the standard *Q*-learning algorithm and observational models are simply extensions of the baseline model of individual learning.

A.1.1. Individual learning

According to the model of individual learning each pair of choices (A and B) is associated with q-values representing the choices' expected consequences at trial t, $Q_{choice}(t)$, initially set to 0. *Q*-values are used with the softmax function for binary choices to calculate the probability of selecting either choice:

$$\rho_A(t) = \frac{\exp(Q_A(t)/\beta)}{\exp(Q_B(t)/\beta) + \exp(Q_A(t)/\beta)} \tag{A.1}$$

where β is the inverse temperature parameter which controls the tendency to explore or exploit data. A low β increases the tendency to exploit data by increasing the probability of selecting the choice with the highest q-value. Following the choice at each trial a prediction error, $\delta_{consequence}$ is calculated as the difference between expected and actual consequence where we set the value of a shock to 1 and the value of not being shocked to −1:

$$\delta_{consequence} = consequence(t) - Q_{selected\ \ choice}(t) \tag{A.2}$$

The prediction error is subsequently used to update the expected consequences of the selected choice using a learning rate, $\alpha_{individual}$:

$$Q_{selected\ \ choice}(t+1) = Q_{selected\ \ choice}(t) + \alpha_{individual} * \delta_{consequence} \tag{A.3}$$

The steps are then repeated for each trial so that the model learns (depending on parameter values) the consequences associated with each choice and selects choices accordingly. For the individual model we set $\alpha_{individual}$ and β to free parameters and since all observational models are extensions of the individual model all models include these two free parameters, see Section A.2. for details on parameter fitting. Note also that this model of individual learning is also used to control the demonstrator's choices for the SD group. The demonstrator's parameters where set $\alpha = 0.3$, $\beta = 0.4$ and the value of the consequences are set to 10 (no shock) and −10 (shock). For the UD group the demonstrator made random choices.

A.1.2. Observational learning

Observational learning is modeled by extending the baseline model of individual learning with learning from observation of choices and observation of consequences respectively.

A.1.2.1. Learning from observation of choices: When the demonstrator's choices are observed during the observation phase an observational choice prediction error, $\delta_{obs.\ choice}$, is calculated as:

$$\delta_{obs.\ choice} = 1 - \rho_{obs.\ choice} \tag{A.4}$$

which is the difference between the probability of the actual choice of the demonstrator (which in hindsight is 1 for the observed choice) and the probability that the model would make the choice given no observable information. Note that the observational choice prediction error is always positive. This prediction error is subsequently used to shift the probabilities to select either choice at the action-phase using an imitation rate conceptually similar to the learning rate in Eq. (A.3):

$$p_{obs.\ choice}(t+0.5) = \rho_{obs.\ choice}(t) + \alpha_{imitation} * \delta_{obs.\ choice} \tag{A.5}$$

$$p_{unobs.\ choice}(t+0.5) = 1 - \rho_{obs.\ choice}(t+0.5) \tag{A.6}$$

where $p_{obs.\ choice}$ and $p_{unobs.\ choice}$ represents the probabilities of making the same choice that the demonstrator made and the opposite choice. For models that included learning from observation of choices we set $\alpha_{imitation}$ to a free parameter. Note that simply extending the model of individual learning described in A.1.1. by including the calculations from Eqs. (A.4), (A.5), (A.6) is equivalent with the CH model.

A.1.2.2. Learning from observation of consequences: When the consequences of the demonstrator's choices are observed, an observational consequence prediction error, $\delta_{obs.\ conseq.}$, is calculated as:

$$\delta_{obs.\ conseq.} = consequence_{obs}(t) - Q_{obs.\ choice}(t) \tag{A.7}$$

where $consequence_{obs}$ represents the consequence following the demonstrator's choice set to 1 or −1 depending on whether or not the demonstrator was presumably given a shock or not. Next, Q-values were updated based on this observable information using an observational learning rate, $\alpha_{obs.\ conseq.}$, similar to updating described in Eq. (A.3):

$$Q_{obs.\ choice}(t+0.5) = Q_{obs.\ choice}(t) + \alpha_{obs.\ conseq.} * \delta_{obs.\ conseq.} \tag{A.8}$$

Subsequently, the probabilities of making either choice at the action stage are calculated as in Eq. (A.1) using the updated Q-value. For models that included learning from observation of consequences we included $\alpha_{obs.\ conseq.}$ as a free parameter. Simply extending the model of individual learning described in A.1.1. by including the calculations from Eqs. (A.7), (A.8) is equivalent with the CO model.

A.1.2.3. Learning from observation of both choices and consequences: We formulated two different models that combined learning from both sources of observable information. The hybrid model (CHCO.H) combined both sources simultaneously and thus used all available information while the separated model (CHCO.S) included observation of choice at the Choice Observation condition but only observation of consequences at the Choice-Consequence Observation condition.

A.1.2.3.1. Hybrid observational learning – CHCO.H: Combining both sources of observable information is carried out such that the probabilities of making either choice at the action stage are calculated in two steps. First, Q-values and subsequent probabilities are calculated as in A.1.2.2. (learning from observation of consequences). Secondly, the probabilities of making either choice are shifted according to the calculations described in A.1.2.1. (learning from observation of choices). This was implemented in the CHCO.H and included setting both $\alpha_{imitation}$ and $\alpha_{obs.\ conseq.}$ to free parameters.

A.1.2.3.2. Separated observational learning – CHCO.S: Separating learning from the two sources of information was simply done by letting model behavior during the Choice-Consequence Observation condition be implemented as in A.1.2.2. (learning from observation of consequences) while behavior at the Choice Observation condition was implemented as in *A.1.2.1. (learning from observation of choices):* This incorporated both sources of information in the model although not simultaneously. This was implemented in the CHCO.S model and included setting both $\alpha_{imitation}$ and $\alpha_{obs.\ conseq.}$ to free parameters.

A.1.2.4. Temporally changing learning from observation of both choices and consequences: The two models that tested whether or not observational learning changed over time where implemented by letting the parameter of interest (either $\alpha_{imitation}$ or $\alpha_{obs.\ conseq.}$) in the CHCO.S model vary over the three blocks of the experiment. This was done by replacing the parameter of interest with two free parameters, one for the first block and one for the last block. For the middle block, the parameter was calculated as lying in between. The models were fitted as before on all trials over all blocks per person. The result can be described as fitting a linearly changing parameter that was kept constant throughout each block. The reason the parameter was kept constant throughout each block was to avoid picking up temporal changes that could be derived from the shift in task that occurred over the course of a block. For instance, learning from observation of consequences might drop over the course of a block when the participant had already learnt the choice-consequence contingency. For these two models based on the CHCO.S model we replaced either the free parameter $\alpha_{imitation}$ with $\alpha_{imitation(F)}$ (first) and $\alpha_{imitation(L)}$ (last) or $\alpha_{obs.\ conseq}$ with $\alpha_{obs.\ conseq(F)}$ and $\alpha_{obs.\ conseq(L)}$.

A.2. Parameter fitting and model comparisons

A.2.1. Parameter fitting

All free parameters were constrained within the interval (0,1) and fitted for each participant over all trials by minimizing the negative log-likelihood, $-ln(L)$, of each model. This was done in R (R Development Core Team, 2012) using the mle2 function from the bbmle package employing the optim optimization function and the BFGS optimization method. To avoid local minima, we fitted

Table A.1
Inclusion of a parameter in a model is marked with an x.

Models	Free parameters							
	$\alpha_{individual}$	β	$\alpha_{imitation}$	$\alpha_{obs.\ conseq.}$	$\alpha_{imitation(F)}$	$\alpha_{imitation(L)}$	$\alpha_{obs.\ conseq.(F)}$	$\alpha_{obs.\ conseq.(L)}$
Individual	x	x	–	–	–	–	–	–
CH	x	x	x	–	–	–	–	–
CO	x	x	–	x	–	–	–	–
CHCO.H	x	x	x	x	–	–	–	–
CHCO.S	x	x	x	x	–	–	–	–
CHCO.t (choice)	x	x	–	x	x	x	–	–
CHCO.t (cons.)	x	x	x	–	–	–	x	x

each set of parameters 40 times with randomized initial parameters and then choose the best fitted parameters. For an overview of which free parameters that were included in each model, see Table A.1.

A.2.2. Model comparisons

Model comparisons were carried out in order to investigate to which extent the different sources of information affected choices and thus did not include the models that looked at the temporal changes. In order to compare the different RL models we calculated the AIC weights, $wAIC_i$, for each model and participant (Wagenmakers, Farrell, & Ratcliff, 2005). AIC weights are calculated using the AIC values that measure the goodness of fit of a model while also taking into account its complexity:

$$\mathrm{AIC} = 2k - 2In \tag{A.9}$$

where k is the number of fitted parameters and $-ln(L)$ is the negative log-likelihood. For each model i ΔAIC_i is calculated as $\Delta AIC_i = AIC_i - AIC_{min}$ were the AIC_{min} is the AIC value for the best model (i.e. the model with the lowest AIC value) for that participant. AIC weights are then calculated by comparing the AIC values for all five models of interest over each participant:

$$wAIC_i = \exp\left(\frac{-\Delta AIC_i}{2}\right) \Big/ \sum_{m=1}^{M} \exp\left(\frac{-\Delta AIC_m}{2}\right) \tag{A.10}$$

where M denotes the number of models compared. AIC weights provide a measure of the weight of evidence for each model in a given set of candidate models and we compared models by looking at the mean and standard deviation of $wAIC_i$, mean rank order and number of wins across participants.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2014.06.010.

References

Apesteguia, J., Huck, S., & Oechssler, J. (2007). Imitation—Theory and experimental evidence. *Journal of Economic Theory, 136*, 217–235.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology, 4*, 1–2.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*(7219), 245–249.

Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science, 33*, 206–242.

Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences of the United States of America, 107*(32), 14431–14436.

Busemeyer, J. R., & Wang, Y. (2000). Model comparisons and model selections based on generalization criterion. *Methodology, 189*, 171–189.

Coolen, I., van Bergen, Y., Day, R. L., & Laland, K. N. (2003). Species difference in adaptive use of public information in sticklebacks. *Proceedings of the Royal Society London B, 270*(1531), 2413–2419.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation and reinforcement learning. *Neuron, 36*, 285–298.

Delgado, M. R., Labouliere, C. D., & Phelps, E. A. (2006). Fear of losing money? Aversive conditioning with secondary reinforcers. *Social Cognitive and Affective Neuroscience, 1*(3), 250–259.

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363*(1511), 3787–3800.

Denrell, J., & Le Mens, G. (2013). Information sampling, conformity and collective mistaken beliefs. In Proceedings of the 35th annual conference of the cognitive science society, 2013 (pp. 2177–2182).

Dewar, G. (2004). Social and asocial cues about new food: Cue reliability influences intake in rats. *Learning & Behavior, 32*(1), 82–89.

Elsner, B., & Aschersleben, G. (2003). Do I get what you get? Learning about the effects of self-performed and observed actions in infancy. *Consciousness and Cognition, 12*(4), 732–751.

Enquist, M., Eriksson, K., & Ghirlanda, S. (2007). Critical social learning: A solution to Rogers's paradox of nonadaptive culture. *American Anthropologist, 109*(4), 727–734.

Feldman, M. W., Aoki, K., & Kumm, J. (1996). Individual versus social learning: Evolutionary analysis in a fluctuating environment. *Anthropological Science, 104*(3), 209–231.

Galef, B. G. (2009). Strategies for social learning: Testing predictions from formal theory. *Advances in the Study of Behavior, 39*, 117–151.

Genz, A., & Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation, 63*, 361–378.

Heyes, C. M., Ray, E. D., Mitchell, C. J., & Nokes, T. (2000). Stimulus enhancement: Controls for social facilitation and local enhancement. *Learning and Motivation, 31*(2), 83–98.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434–446.

Kameda, T., & Nakanishi, D. (2003). Does social/cultural learning increase human adaptability? Rogers's question revisited. *Evolution and Human Behavior, 24*(4), 242–260.

Kavaliers, M., Choleris, E., & Colwell, D. D. (2001). Learning from others to cope with biting flies: social learning and fear-induced conditioned analgesia and active avoidance. *Behavioral Neuroscience, 115*(3), 661–674.

Kendal, R. L. (2004). The role of conformity in foraging when personal and social information conflict. *Behavioral Ecology, 15*(2), 269–277.

Kendal, R. L., Coolen, I., van Bergen, Y., & Laland, K. N. (2005). Trade-offs in the adaptive use of social and asocial learning. *Advances in the Study of Behavior, 35*, 333–379.

Kendal, J. R., Rendell, L., Pike, T. W., & Laland, K. N. (2009). Nine-spined sticklebacks deploy a hill-climbing social learning strategy. *Behavioral Ecology, 20*(2), 238–244.

Laland, K. N. (2004). Social learning strategies. *Learning & Behavior, 32*(1), 4–14.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice* (vol. 2010). SAGE Publications, Inc.

McElreath, R., Bell, A. V., Efferson, C., Lubell, M., Richerson, P. J., & Waring, T. (2008). Beyond existence and aiming outside the laboratory: Estimating frequency-dependent and pay-off-biased social learning strategies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363*(1509), 3515–3528.

McElreath, R., Fasolo, B., & Wallin, A. (2013). The evolutionary rationality of social learning. In R. Hertwig & U. Hoffrage (Eds.), *Simple heuristics in a social world*. Oxford University Press.

McGregor, A., Saggerson, A., Pearce, J., & Heyes, C. (2006). Blind imitation in pigeons, Columba livia. *Animal Behaviour, 72*(2), 287–296.

Merlo, A., & Schotter, A. (2003). Learning by not doing: An experimental investigation of observational learning. *Games and Economoc Behavior, 42*, 116–136.

Mesoudi, A. (2008). An experimental simulation of the "copy-successful-individuals" cultural learning strategy: Adaptive landscapes, producer–scrounger dynamics, and informational access costs. *Evolution and Human Behavior, 29*(5), 350–363.

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society B, 279*, 653–662.

Nicolle, A., Symmonds, M., & Dolan, R. J. (2011). Optimistic biases in observational learning of value. *Cognition, 119*(3), 394–402.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology, 14*(6), 769–776.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience, 2*(1), 3–11.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10*(9), 1095–1102.

R Development Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from <www.R-project.org>.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Laland, K. N. (2010). Why copy others? Insights from the social learning strategies tournament. Science (New York, N.Y.), 328(5975), pp. 208–13.

Rescorla, R. A. (1969). Establishment of a positive reinforcer through contrast with shock. *Journal of Comparative and Physiological Psychology, 67*(2), 260–263.

Rushworth, M. F. S., Mars, R. B., & Summerfield, C. (2009). General mechanisms for making decisions? *Current Opinion in Neurobiology, 19*(1), 75–83.

Schlag, K. H. (1999). Which one should I imitate? *Journal of Mathematical Economics, 31*(4), 493–522.

Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making, 17*(3), 173–190.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88*(2), 135–170.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, Massachusetts: The MIT Press.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., & Haruno, M. (2012). Learning to simulate others' decisions. *Neuron, 74*(6), 1125–1137.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology. General, 134*(1), 108–116.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. England: University of Cambridge.

Webster, M. M., & Laland, K. N. (2008). Social learning strategies and predation risk: Minnows copy only when using private information would be costly. *Proceedings. Biological Sciences/The Royal Society, 275*(1653), 2869–2876.