Computational Models for the Combination of Advice and Individual Learning

Guido Biele,[a] Jörg Rieskamp,[a,b] Richard Gonzalez[c]

[a]*Max Planck Institute for Human Development*
[b]*Department of Psychology, University of Basel*
[c]*Department of Psychology, University of Michigan, Ann Arbor*

Abstract

Decision making often takes place in social environments where other actors influence individuals' decisions. The present article examines how advice affects individual learning. Five social learning models combining advice and individual learning-four based on reinforcement learning and one on Bayesian learning-and one individual learning model are tested against each other. In two experiments, some participants received good or bad advice prior to a repeated multioption choice task. Receivers of advice adhered to the advice, so that good advice improved performance. The social learning models described the observed learning processes better than the individual learning model. Of the models tested, the best social learning model assumes that outcomes from recommended options are more positively evaluated than outcomes from nonrecommended options. This model correctly predicted that receivers first adhere to advice, then explore other options, and finally return to the recommended option. The model also predicted accurately that good advice has a stronger impact on learning than bad advice. One-time advice can have a long-lasting influence on learning by changing the subjective evaluation of outcomes of recommended options.

Keywords: Reinforcement learning; Social learning; Advice taking; Learning model; Decision making

Many decisions are made in a social context, where decision makers can observe others' decisions or receive advice from other people. Accordingly, it has frequently been argued that we learn how to make decisions from others (e.g., Bandura, 1977; Henrich & McElreath, 2003; Laland, 2001; Schotter & Sopher, 2003; Simon, 1990). Social information seems especially valuable in situations of uncertainty (Festinger, 1954), for instance, when

Correspondence should be sent to Guido Biele, Max Planck Institute for Human Development, Lentzealle 94, 14195 Berlin, Germany. E-mail: biele@mpib-berlin.mpg.de

decision makers have little knowledge about the judgment domain, when choice options' outcomes seem similar, or when information about choice options needs to be collected through first-hand experience.

In many real-life situations people learn to make better decisions based on the experienced consequences of their own decisions (Barron & Erev, 2003; Busemeyer & Myung, 1992; Denrell, 2005; Hertwig, Barron, Weber, & Erev, 2004; March, 1996). For example, decision makers learn inference strategies to judge companies' credit worthiness (e.g., Rieskamp & Otto, 2006) or strategies for social interaction (e.g., Stahl, 1996), consumers choose detergents based on their experience with different brands (e.g., Erdem & Keane, 1996), and through experience investors learn how to allocate financial resources (e.g., Goetzmann & Massa, 2002; Rieskamp, 2006a; Rieskamp, Busemeyer, & Laine, 2003).

Empirical evidence suggests that decision makers do not rely exclusively on their own experience in these situations, but they also learn from others. For instance, employees learn how to make decisions from their colleagues (Gibson, 2004), consumers can get advice from friends or publications such as *Consumer Reports*; and investors are influenced by other investors' decisions (e.g., Roider, Drehmann, & Oechssler, 2007). In the present article, we argue that in decision situations characterized by uncertainty and incomplete knowledge, advice strongly influences people's learning processes. Specifically, we examine how one-time advice from another person influences individual learning processes in repeated choice tasks.

In the next section, we briefly review research on individual and social learning in decision making. Then we introduce the experimental paradigm we used and report whether decision makers follow advice (Experiment 1). We present the learning models and examine how well they describe behavior of the receivers of advice (henceforth ''receivers''). Experiment 2 was designed specifically to test the models that best explained the data from Experiment 1. We conclude with a general discussion.

1. Experienced-based decision making and learning

In a typical decision-from-experience task, the decision maker repeatedly chooses between two or more options with unknown expected payoffs (e.g., Hertwig et al., 2004; see also Shanks, Tunney, & McCarthy, 2002). How people deal with such decision situations under uncertainty has been studied extensively with various tasks (e.g., Erev & Barron, 2005; Estes, 1962; Gans, Knox, & Croson, 2007; Hutchinson & Meyer, 1994; Meyer & Shi, 1995; Murray, 1971; Vulkan, 2000), and a number of computational models have been proposed to described people's decisions and learning processes. For instance, Busemeyer and Myung (1992) described how decision thresholds for categorization decisions change through experience. Likewise, Busemeyer and Stout (2002) showed that a simple reinforcement learning model describes learning better than a Bayesian learning model in the Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994), in which people learn to choose the best out of four risky options. Gans et al. (2007) explored the predictive

accuracy of various learning models for a two-armed-bandit task with repeated choice between two options. They showed that an exponential smoothing model, which exponentially weighs past experiences to determine current expectations, best predicted the repeated choices. Erev and Barron (2005) (see also Rieskamp, 2006b; Rieskamp & Otto, 2006) proposed a learning model which assumes that people learn to choose among cognitive strategies. Yechiam and Busemeyer (2005) examined the various assumptions of learning models, for example, they compared learning models' various updating rules for integrating new experiences with accumulated past experience. They found that learning models that assume that choice propensities decay as a function of time alone, independently of whether an option was chosen or not, explained the data better than models which assume that forgetting old experience is limited to the currently updated option. In sum, this research suggests that decision making from experience can be aptly described by simple learning models that assume people form choice propensities based on exponentially discounted experience, and that they then decide based on these propensities.

2. Social learning and decision making

A feature common to all the learning models mentioned above is the assumption that learning is based exclusively on individual experience. In contrast, theories of social learning describe how social information influences people's behavior. Bandura's (1977) prominent social learning theory assumes that people learn—simple behavior as well as complex concepts—by observation and cognitive modeling (Rosenthal & Zimmerman, 1978; Zimmerman & Rosenthal, 1974).

Past research on social learning has examined the impact of social interaction or face-to-face interaction on learning. Recent research has followed a broader definition of social learning that includes decision making and learning processes influenced by social *information* gathered from others, and which was not necessarily acquired through any personal interaction. We follow this broader definition of social learning. Recent social learning research also places stronger emphasis on computational models of social learning. For instance, inspired by Boyd and Richerson's (1985) theory of gene–culture coevolution, McElreath et al. (2005) proposed a model of imitation learning that combines individual learning with social learning by assuming that a choice option is reinforced through received payoffs and through the observation that others choose that option. Apesteguia, Huck, and Oechssler (2007) examined imitation behavior in repeated social interactions. They reported that the probability of imitating another person increases with the payoff difference between the learner and the other person, so that more successful players are more likely to be imitated (see also Schlag, 1998, 1999).

Another line of research examines how individuals seek and integrate advice when making nonrepeated decisions. Luan, Sorkin, and Itzkowitz (2004) examined the influence of advice in signal detection tasks and found that decision makers are sensitive to the quality of advice—they give more weight to better advisors—and they search for advice in an adaptive manner when they can decide whether, and from whom, to seek advice. Budescu and

Rantilla (2000) described how decision makers integrate expert opinions with a model that weights experts' advice according to the amount of information that advisors had available to generate advice. Yaniv and Kleinberger (2000) found that decision makers gave too little weight to others' advice, and Yaniv (2004b) reported that more-knowledgeable decision makers discount advice more frequently than less-knowledgeable decision makers. Importantly, Yaniv (2004a,b) pointed out that advice from independent decision makers generally improves performance. In sum, while the experimental tasks and objectives of research on social learning are diverse, a common finding is that people do not rely exclusively on either their own judgment or on received advice, but they combine both.

We provide a complement to existing models of advice taking by explaining how advice affects individual learning. Earlier models of advice taking have focused on single decisions and not on repeated choices, as we do. The work on imitation learning has examined people's behavior when others' behavior can be observed on a trial-by-trial basis. In contrast, we focus on a situation in which the decision maker cannot observe and imitate others' behaviors but solves the task in isolation after receiving an initial piece of advice. To describe how the individual learning process is affected by advice we propose and test several learning models. Our social learning models are built upon simple learning models because these models have been successful in describing people's decisions in repeated choice tasks, as described above.

3. Experiment 1

Experiment 1 examined how social learning influences choice. Participants individually solved a repeated choice task in which they could learn to choose the best alternative. Additionally the participants received advice on how to solve this task. In the experiment, the Iowa Gambling Task (Bechara et al., 1994) was presented to participants. The IGT is a type of armed bandit task in which participants try to obtain high rewards by repeatedly choosing the best of four choice options, which are associated with different payoff distributions. Participants receive feedback about the outcomes of their choices to learn which of the several options provides the highest average payoff. The challenge of the IGT lies in the options' payoff distributions. The options that lead on average to the highest payoffs frequently lead to low gains but do not encounter high losses. In contrast, options that lead on average to low payoffs frequently lead to larger gains but lead to even higher losses.

In our experiment the participants received an initial endowment (10 euros) and then chose cards from four card decks (A, B, C, D). When a participant chose from deck A or B, he or she always received a reward of 50 eurocents; when the choice was from deck C or D, the participant always received 25 eurocents. Importantly, participants sometimes additionally incurred a loss when choosing from a deck. Losses when choosing from decks C and D (henceforth ''good decks'') were moderate, so that the expected payoff from those decks after 100 trials was 12.5 euros. Losses from decks A and B (henceforth ''bad decks'') were so large that the expected payoff from these decks was –12.5 euros. The difference between decks with the same expected payoff was that one deck had frequent but lower losses (low

variance), whereas the other deck had rare but higher losses (high variance). The applied payoff schedule was identical to the schedule introduced by Bechara et al. (1994), which has the property that the decks' average payoffs are maintained for blocks of 10 choices. A crucial property of this schedule is that losses from the bad decks occur relatively late, so that the bad decks initially seem to be better. Participants usually need at least 20 trials to learn which decks are best, and after that they still frequently choose one of the bad decks (Maia & McClelland, 2004).

Experiment 1 examines whether social learning can improve decision makers' performance by helping them detect the good decks earlier and also by increasing their likelihood of choosing good decks later in the task. An important property of the IGT is that the two good decks have identical expected payoffs, so that adherence to advice can be tested by examining how frequently participants stay with the recommended deck in the presence of an equally attractive alternative (henceforth ''corresponding deck'').

3.1. Design

To examine the effect of social learning, participants performed a computerized version of the IGT with and without advice. Participants in the independent condition performed the task without receiving or giving advice. Participants in the advisor condition performed the IGT without receiving advice, then chose one of several predetermined advice strategies for another participant, and finally performed the IGT again. Participants in the receiver condition received advice from an advisor and then performed the IGT.[1]

3.2. Participants and procedure

Ninety participants, mostly students from the Free University of Berlin (54% women; mean age of 25 years), were randomly assigned to the three conditions. The experiment was conducted in sessions with two to six participants. In the independent condition, participants were instructed that they were taking part in a decision-making experiment in which they would repeatedly choose cards from four card decks. It was then explained that drawing a card would always lead to a gain or a loss, which would be depicted on the back of the card, and that the gain or loss would be added to their account. The instructions also explained that one could learn during the experiment which payoffs were associated with which decks.

To inform participants about the stochastic nature of the task, it was explained that the payoffs from the card decks were determined before the experiment began, and that participants' choices could not influence the decks' payoffs or the order of one deck's payoffs. To further clarify the stochastic nature, the last 20 participants in each condition were asked to imagine that they were choosing from actual card decks. The instructions included no statement about possible time dependencies of the payoff distributions. Behavior (i.e., frequency of choosing good decks and adherence to advice by receivers) was similar for all participants, so we will not distinguish between the first 10 and last 20 participants in the conditions.
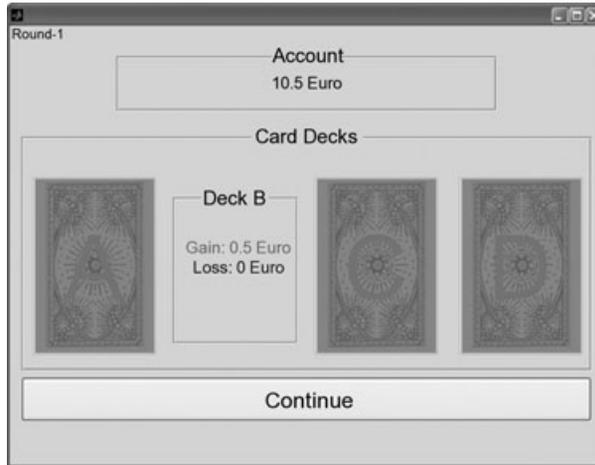
Fig. 1. Graphical user interface for participants in Experiments 1 and 2. Participants chose decks by clicking on the decks. After each choice, feedback was presented, with gains in green and losses in red font (the original user interface was in German).

After the introduction of the task, participants were told that they would start the task with an initial endowment of 10 euros. They were reminded that they would receive their final account balance minus the 10-euro initial endowment as a variable payment. In addition to performance-contingent payment, all participants received a show-up payment of 5 euros. In the rare case that the final account balance minus the 10-euro initial endowment was negative, participants still received the show-up payment of 5 euros, but they only learned this after the experiment. Finally, participants were briefly instructed about the graphical user interface (see Fig. 1) used to conduct the experiment. After choosing from a deck by clicking on it, the display showed participants the gain (in green) and the loss (in red) associated with the card. At the same time, the overall account was updated with the payoff of the current choice. Participants clicked the ''continue'' button to go on to the next trial. The minimum time interval between two choices was fixed to 3 s; no upper time limit was set.

Advisors received the same information as the independent decision makers, plus additional information about their role as advisors. Specifically, they were first informed that they would advise another participant who would perform the identical task. To be able to evaluate whether receivers actually followed the advice, a set of four feasible recommendations was predefined, namely, ''always choose from deck A'' (or ''B,'' or ''C,'' or ''D''). The feasible advice was presented to advisors before they made their first 100 choices. Advisors were not informed that they would encounter the same task again after giving the advice (henceforth the second 100 choices). Advisor and receiver always participated in the same session. To communicate advice, an advisor marked his or her advice with a pen on a form, which was then given to the receiver. To motivate advisors and to make them credible to receivers, they received an amount equal to 50% of the receiver's payoff, in addition to the payoff from their own choices.

Receivers were first provided with the instructions for the IGT and then with a form where one of the feasible lines of advice was marked by an advisor in the same session. Receivers were aware that the advisor was in the same experimental session. To prevent personal communication between advisor and receiver, they were seated in separate cubicles. To clarify the experience and motivation of advisors, receivers were informed that the advisor had participated in the same task prior to choosing his or her recommendation, and that the advisor would receive a payment equivalent to 50% of the receiver's payment from the IGT. As in the other conditions, receivers' payments varied depending on their performance in the IGT.

3.3. Results

3.3.1. Choices and performance

Participants earned, on average, 5.02 euros (SD = 5.31) in the IGT. Independent decision makers chose one of the two good decks, on average, in 62% (SD = 14%) of the 100 trials, which is less than the proportion of 78% (SD = 2%) in which advisors chose one of the good decks in their last 100 trials, $t(29)$ = 4.13, p < .001, d = 1.06, and less than receivers with 73% (SD = 17), $t(29)$ = 2.54, p < .001, d = .66. Receivers chose one of the good decks across their 100 trials more frequently than advisors in their first 100 trials, $t(29)$ = 3.11, p = .003, d = .8. The advisors chose one of the good decks in 59% (SD = 14) of the first 100 trials and in 78% of the last 100 trials; thus, they improved their performance significantly from the first to the second block of 100 trials, $t(29)$ = 4.75, p < .001, d = 1.23.

Fig. 2 shows, in blocks of 10 trials each, the proportion of participants who chose one of the two good decks. This proportion declined at the beginning (i.e., 10–20 trials) for all groups, with the exception of the advisors at the beginning of their second 100 trials (i.e., postadvice giving). Fig. 2 also shows that in the first 10 trials receivers performed better than advisors in the beginning of their second 100 trials. However, starting at about trial 15, the receivers performed worse than the advisors, and only at the end of the 100 trials did both groups perform equally well again. In sum, advice generally improved performance, compared to inexperienced participants, with the advantage being especially large in the early trials.

3.3.2. Giving and following advice

A large majority of participants (28; 93%) in the role of advisors gave good advice. Of these 28 advisors, 19 proposed the good deck with a high payoff variance (rare but high losses) and nine proposed the good deck with a low payoff variance (frequent but low losses). Advisors chose the deck they recommended on average 42% (SD = 19%) of the first 100 trials, indicating that they recommended their preferred deck.

To examine the influence of advice we tested whether receivers chose the recommended deck (regardless of whether the advice was good or bad) more frequently than the corresponding deck with the same expected payoff. Receivers chose the good deck with low variance when it was recommended, on average, in 62% of the trials (SD = 9%), whereas the mean percentage was 10% (SD = 3%) when it was not recommended. The mean percentage
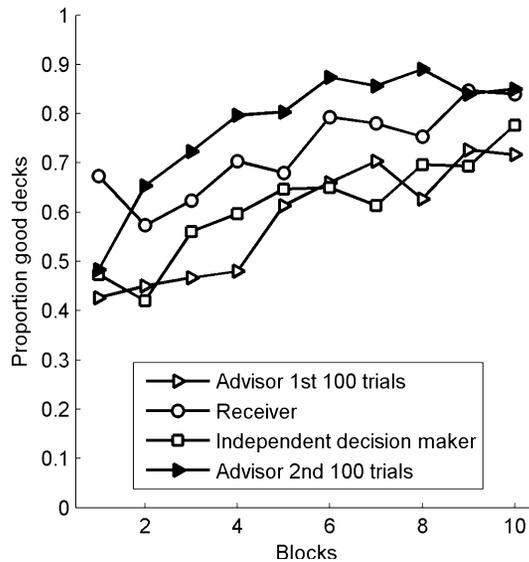
Fig. 2. Participants' average proportion of choosing one of the good decks in Experiment 1 (in blocks of 10 trials). Proportions were first calculated for each participant and block and then averaged. Receivers perform constantly better than individual learners and advisors in their first 100 trials, but worse than advisors in their second 100 trials.

for the good deck with high variance was 69% (SD = 9%) when it was recommended but only 7% (SD = 3%) otherwise. Fig. 3 shows the development of choice proportions over the 100 trials and reveals that adherence to advice declined from the first block (all participants chose the recommended deck in the first round) to the second block and then rebounded to high adherence rates again. We refer to this sequence as the adherence–exploration–adherence pattern. Altogether, the results show a strong influence of advice on choices because receivers clearly preferred the recommended deck to the corresponding deck with the same expected payoff.

The analyses show that participants without advice learned to choose the good decks, receivers followed the advice they received, and having received advice gave the receivers an advantage—especially in the early choices. The decline and rebound of the probability with which receivers chose according to the advice suggests that they combined recommendations with individual experience to determine which choice to make.

4. Models of learning in repeated choice tasks

As reinforcement learning models have been most successful in describing people's choices in instrumental learning situations (Busemeyer & Stout, 2002; Bush & Mosteller, 1955; Erev & Barron, 2005; Estes, 1950; Gans et al., 2007; Yechiam & Busemeyer, 2005), we concentrate our examination on variants of these models. The learning models we propose are similar to the models suggested by Erev and Roth for learning in experimental
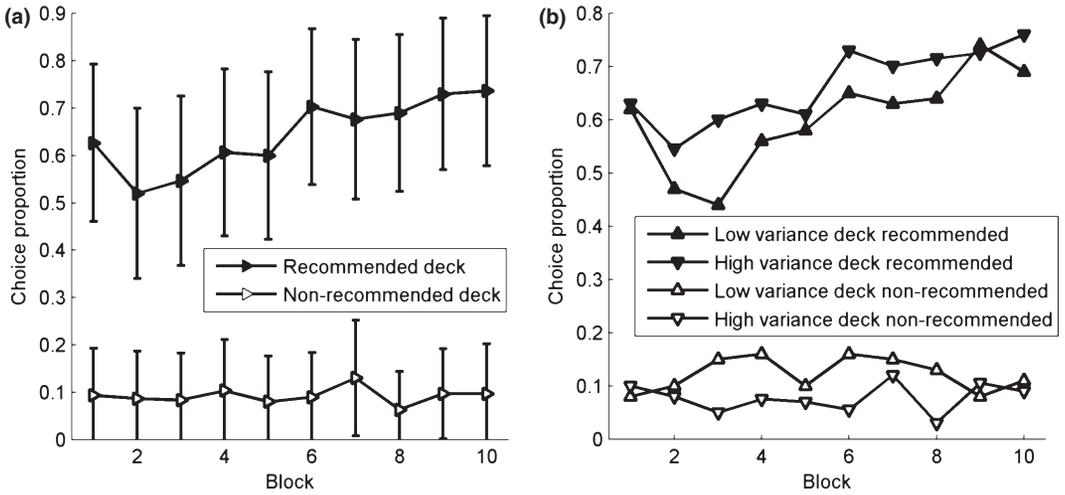
Fig. 3. Participants' choice proportions conditional on advice in Experiment 1 (blocks of 10 trials). The left panel (a) shows preferences for the recommended deck (▲) versus the preference for the corresponding deck with the same expected payoff (△). Error bars are 1.96 times the standard error of the mean. The right panel (b) distinguishes between choices after the recommendations to choose a low variance deck versus the recommendation to choose a high variance deck. Choice proportions were first calculated for each participant and block and then averaged. The figure shows that the recommended deck was always favored, regardless of whether advice was to choose a deck with high or low variance. Note that decks with the same expected payoff do not need to sum up to one because participants can also choose from the two other decks.

games (Erev, 1998; Erev & Roth, 1998) and by Busemeyer and colleagues for learning in the IGT (Busemeyer & Stout, 2002; Yechiam & Busemeyer, 2005).

In the reinforcement learning task the decision maker repeatedly chooses an option i from a set $S = \{1, 2, ..., n\}$ with n options. Before making a decision, the decision maker might receive advice to choose an option or options $A = \{1, ..., m\}$, where A is a subset of S. Generally, advice can consist of one or several options. After choosing option i in trial t, the decision maker receives a payoff $\pi_t(i)$.

4.1. Individual learning

According to the individual reinforcement learning (RL) model proposed here, the decision maker enters the situation with initial propensities to choose the different options. After choosing an option, the resulting payoff is used to update the option's choice propensity. Independent of choice, the propensities decay with time. Choice probabilities are an increasing function of the options' propensities. Formally, the initial propensity of an option is $q_1(i) = 0$. After choosing an option i, the propensities $q(i)$ of the options are updated by

$$q_{t+1}(i) = (1 - \phi) \cdot q_t(i) + r_t(i), \tag{1}$$

where ϕ is a free decay parameter determining the weight of past experiences in the updating process, with $r_t(i) = \pi_t(i)$ for the chosen option and $r_t(i) = 0$ for options not chosen.

The probability of choosing an option is defined by

$$p_t(i) = e^{\lambda \cdot q_t(i)} / \sum_{j=1}^{n} e^{\lambda \cdot q_t(j)}. \tag{2}$$

To capture the variability in participants' sensitivity to differences in propensities, the choice rule is augmented by a sensitivity parameter λ (e.g., Yechiam & Busemeyer, 2005). We further assume that participants who received advice will choose the recommended option in their first trial (28 out of 30 participants in Experiment 1 behaved accordingly). This makes RL a nested version of the more complex social learning models described next and also makes the RL model a strong competitor of the social learning models.[2]

4.2. Social learning

The following social learning models combine information received as advice with an individual reinforcement learning process and are, therefore, called advice-reinforcement combination (ARC) models. To specify the ARC models formally, we modify the individual RL model described above by adding mechanisms to it. Because we assume that receivers will attempt to evaluate the recommended option before exploring alternative options, all tested models choose the recommended option in the first trial. Formally, the probability of choosing an option in the first trial is $p(i|i \in A) = 1/m$ and $p(i|i \notin A) = 0$.

4.2.1. ARC-Initial
One way to introduce social information into the individual learning process is to assume that decision makers initially perceive recommended options as more positive than non-recommended options. This assumption is reasonable because advisors usually have more knowledge than receivers (Jungermann & Fischer, 2005). To model the initial preference for the recommended option, we allowed the initial propensity of the recommended option to be higher than for options that were not recommended. Similarly, Camerer, Ho, and Chong (2002) and Hanaki, Sethi, Erev, and Peterhansl (2005) modeled a decision maker's own past experience by defining initial propensities as a function of the options' past payoffs. Formally, the initial propensities for ARC-Initial are defined as $q_1(i|i \in A) = |\mu| \cdot \iota$ and $q_1(i|i \notin A) = 0$, where ι is a free parameter determining the extra initial propensity of the recommended option, and μ is the expected payoff from always choosing the best option, which allows the interpretation of ι independent of the specific payoff distribution.

4.2.2. ARC-Outcome-bonus
Social information could also influence the ongoing evaluation of payoffs. The outcome-bonus model assumes that the consequences of recommended options are perceived more positively, compared to the consequences of nonrecommended options. This assumption is consistent with research on imitation by Miller and Dollard (1941), which shows that imitation in itself can become a secondary reinforcer. Alternatively, one could consider following

advice to be cooperative behavior, which can also be intrinsically rewarding (Decety, Jackson, Sommerville, Chaminade, & Meltzoff, 2004).

A generally more positive evaluation of outcomes from recommended options can be implemented by adding a constant bonus to every payoff from the recommended option. Formally, reinforcements for recommended options are $r_t(i|i \in A) = \pi_t(i) + |\mu| \cdot \rho$, where ρ is a free parameter specifying the additional reinforcement for choosing a recommended option.

4.2.3. ARC-Decay

It can be assumed that recommended options have, due to their prominence, stronger memory traces and are therefore easier to retrieve (Lockhart, 2001). Hence, it should be easier to retrieve information about the past performance of a recommended option. We implement this assumption by introducing an additional decay parameter ϕ_{advice} for the recommended option, which is assumed to be lower than the decay parameter ϕ for all other options. The important implication of this model is that the accumulation of (negative or positive) propensities of the recommended option will be faster, and their reduction as a function of time slower. Formally, the different decay process is implemented by modifying Equation 1 to

$$q_{t+1}(i) = \begin{cases} i \notin A \rightarrow (1 - \phi) \cdot q_t(i) + r_t(i) \\ i \in A \rightarrow (1 - \phi_{advice}) \cdot q_t(i) + r_t(i) \end{cases}. \tag{3}$$

4.2.4. ARC-Certainty

The social learning models presented thus far assume that social information directly influences the learning mechanism. Alternatively, as proposed by Festinger (1954), people might generally rely on individual learning and resort to advice only when uncertain about how to evaluate the available options (see also Henrich & Boyd, 1998; Kameda & Nakanishi, 2002, 2003). We implement reliance on social information in situations of uncertainty by assuming that decision makers choose according to propensities when the variance of propensities is high and choose the recommended option when the variance of propensities is low. To make the social learning parameter independent of payoff magnitudes, reliance on advice is modeled as contingent on the variance of the choice probabilities. Specifically, the choice probabilities are modified, after they are determined by Equation 2, according to the following function:

$$p(i) = \begin{cases} \sigma(\Pi) < \tau \wedge i \in A \rightarrow 1/m \\ \sigma(\Pi) < \tau \wedge i \notin A \rightarrow 0 \\ \sigma(\Pi) \geq \tau \rightarrow p \end{cases}, \tag{4}$$

where $\sigma(\Pi)$ is the standard deviation of the choice probabilities in Equation 2 and τ is a free parameter that determines the threshold below which the recommended option is chosen.[3]

4.2.5. Bayesian advice integration

The learning models described above are based on learning models that have successfully explained learning in previous experiments. In contrast to a Bayesian

learning approach they give more weight to recent experiences. It is possible that in the current experiment, which explicitly demands the integration of the prior information of advice with new evidence, a Bayesian approach could provide a good account of people's behavior. To investigate this possibility, we also test a Bayesian model that was explicitly suggested for the IGT by Yechiam and Busemeyer (2005). This model uses the following Bayesian updating rule:

$$q_{t+1}(i) = \frac{\phi_B + N_{t-1}(i)}{\phi_B + N_t(i)} \cdot q_t(i) + \frac{1}{\phi_B + N_t(i)} \cdot r_t(i),$$ (5)

where $N_t(i)$ is the total number of times option i was chosen up to round t, and ϕ_B is the weight given to initial expectations. The influence of advice on learning is incorporated in the initial evaluation of the choice options, which is, as in ARC-Initial, assumed to be higher for the recommended deck, $q_1(i|i \in A) = |\mu| \cdot \iota$ and $q_1(i|i \notin A) = 0$. The Bayesian model uses the same choice rule (Equation 2) as all other models.

4.3. Comparison of social learning mechanisms

Table 1 summarizes the functions used to describe the ARC models and the RL model. To highlight the differences between the models we examine qualitative and quantitative aspects of the models' predictions about learning:

ARC-Initial assumes that advice mainly affects the initial learning process due to the changed initial propensities of the recommended options. The decay of propensities implies that the influence of advice can rapidly be negligible (depending on the rate of decay). As a result, receivers will also learn over time to deviate from a recommended

Table 1
Differences of the learning and choice mechanisms for the individual reinforcement learning (RL) model and the advice-reinforcement combination (ARC) models: ARC-Initial, ARC-Outcome-bonus, ARC-Decay, and ARC-Certainty

Mechanism	RL Model	ARC Models
First choice	$p(i|i \in A) = 1/m$ and $p(i|i \notin A) = 0$	No difference
Initial attraction	$q_1(i) = 0$	ARC-Initial: $q_1(i|i \in A) = |\mu| \cdot \iota$ and $q_1(i|i \notin A) = 0$
Reinforcement	$r_t(i) = \pi_t(i)$	ARC-Outcome-bonus: $r_t(i) = \begin{cases} i \in A \to \pi_t(i) + |\mu| \cdot \rho \\ i \notin A \to \pi_t(i) \end{cases}$.
Updating	$q_{t+1}(i) = (1 - \phi) \cdot q_t(i) + r_t(i)$	ARC-Decay: $q_t(i) = \begin{cases} i \in A \to q_t(i) \cdot (1 - \phi_{advice}) + r_t \\ i \notin A \to q_t(i) \cdot (1 - \phi) + r_t \end{cases}$.
Choice rule	$p(i) = e^{\lambda \cdot q(i)} / \sum_{j=1}^{n} e^{\lambda \cdot q(j)}$	ARC-Certainty: $p(i) = \begin{cases} \sigma(\Pi) < \tau \wedge i \in A \to 1/m \\ \sigma(\Pi) < \tau \wedge i \notin A \to 0 \\ \sigma(\Pi) \geq \tau \to p \end{cases}$.

Note: The second column describes the mechanisms in the individual learning model. The third column shows how the RL model is modified for the respective model to incorporate social learning.

option when a better alternative is available. The impact of advice depends mainly on the magnitude of the initial propensity for the recommended option as well as the decay parameter. The declining influence of advice also implies that ARC-Initial should have difficulty describing the adherence–exploration–adherence pattern. ARC-Initial influences the learning process in favor of the recommended option in both the gain and the loss domains.

According to the outcome-bonus model, the influence of social information accumulates during learning so that its impact is relatively small at first but increases thereafter. Therefore the outcome-bonus model can predict the adherence–exploration–adherence pattern. The model also predicts that the choice of the recommended option increases in the gain and loss domain. With regard to bad advice the outcome-bonus model predicts constant influence on the learning process because the additional reinforcement for the wrongly recommended option makes it continue to appear better than it actually is. This also implies that when a receiver compares two good options with an identical expected value, then the recommended option will subjectively be perceived more positively, that is, the receiver will not become indifferent to the equally good options.

According to the decay model, advice has a constant influence on the learning process and can explain why receivers come to prefer the recommended option again after an exploratory phase and why in the gain domain receivers adhere to the recommended option in the presence of better alternatives. In the loss domain, the decay model predicts that receivers will tend to avoid recommended options because the slower decay for propensities of this option will maintain negative propensities longer, thus strengthening the advantage of alternative options.

Social learning according to ARC-Certainty depends on the choice options' similarity, but not on the amount of individual experience. Generally, the model predicts strong social influence when the choice options' expected outcomes do not differ substantially. ARC-Certainty can predict a preference for the recommended option over a better alternative in cases where the variance across options' expected outcomes is low, or when payoff differences are not perceived due to high within-option payoff variances.

In the long run, the Bayesian model of advice integration makes similar predictions as ARC-Initial. The important differences are that due to the decay process, ARC-Initial shows higher sensitivity to recent outcomes and the exponential decay in ARC-Initial means that the influence of advice diminishes more quickly than in the Bayesian model.

In sum, the comparison of the social learning models shows that they make different predictions independently of specific parameter values. ARC-Initial predicts that the influence of advice should be particularly strong at the beginning of the learning phase. In contrast, the other models—in particular the outcome-bonus model—assume that advice is still effective in later stages of the learning process and are thus better able to explain a robust effect of bad advice and the adherence–exploration–adherence pattern. Of the models, only the decay model is consistent with faster deviation from advice when the expected payoff from the recommended option is negative, independently of whether the advice was good or bad.

5. Evaluating models of social learning

5.1. Parameter estimation

The first step needed to evaluate the learning models was to estimate their parameters. We estimated the models' parameters for each participant separately. We think it is important to model behavior at the individual level, because false conclusions about the distribution of model parameters can be drawn when considering only aggregate data (Estes & Maddox, 2005). A second decision concerned the question of whether the predictions of a model for a person should rely on the past behavior of that person. When using the one-step-ahead approach, the propensities of a learning model for a new trial are updated based on the real choices and payoffs a participant received in the preceding trials. In contrast, when using the simulation approach, the propensities are updated based on the payoffs of the predicted choices. The one-step-ahead approach tends to fit the learning process better because incorrect predictions do not enter the updating process and the model can therefore describe a broader range of behaviors.[4] For this reason, we chose the more demanding simulation approach to estimate the models' parameters, where real choices are only used to determine the model fit but not to determine choices, thus providing a more illuminating test of the model.

For each trial, all models determine the probability with which an individual will choose any option based on past choices and parameter values. We relied on maximum likelihood estimation to find the best parameter values; that is, we searched for the parameter values that maximized the sum of the log likelihood of the observed choices of the four decks. The sum of the log likelihood is defined as $\text{LL} = \sum_{t=1}^{T} \ln p_t(k)$, with T as the number of trials and $p_t(k)$ as the probability with which the model predicts the actual choice k of the participant in trial t. As the logarithm of zero is minus infinity, we fixed the minimum choice probabilities in the fitting process to .001. To determine a choice for each trial, one of the options was randomly selected according to the model's predicted choice probabilities. Due to this random element the model's predictions for a particular set of parameter values were simulated 50 times, and the average learning process of all 50 simulations was determined. The likelihood of the data was then determined based on the average choice probabilities over the 50 simulations of the complete learning process.

The model parameters were constrained to $\phi \in [0,1]$ (for Bayesian updating $\phi_B \in [0,100]$) and $\delta \in [0,1]$ for the decay parameters, $\lambda \in [-5,5]$ for the sensitivity parameter, $\rho \in [0,10]$ for the additional reinforcement in ARC-Outcome-bonus, $\iota \in [0,100]$ for the higher initial attraction in ARC-Initial, and $\tau \in [0,.5]$ for the threshold in ARC-Certainty, where .5 is the maximum standard deviation over choice probabilities for a choice set with four options (the best-fitting parameters did not approach the boundaries, with the exception of the sensitivity parameter of the individual learning model for one participant). To identify the best parameter values, we first performed a grid search and then used the five best parameter sets from the search as starting values for the simplex optimization algorithm (Nelder & Mead, 1965) to determine the best fitting parameter values.

5.2. Model comparison

To evaluate model performance, each model was first compared to a statistical baseline model with three parameters that assumed decision makers always choose according to constant choice probabilities, which were determined by the proportion of how often specific decks were chosen by the participants over the 100 trials (e.g., when a participant had chosen deck A 50 times the baseline model predicted its choice with a constant probability of .50). Because the baseline model was fitted to the data, it is a strong competitor to the learning models. A learning model will only do better than the baseline model if it accurately describes how people change their behavior over time. To account for differences in model complexity we used the Bayesian information criterion (BIC, see, e.g., Zucchini, 2000) as a model selection criterion. The BIC is defined as $-2 \times \text{LL(model)} + \text{number of parameters} \times \log(N)$, with N as the number of predicted choices, and leads to positive values with smaller values indicating a better fit. We compared each model with the statistical baseline model by determining the baseline model's BIC value and subtracted the BIC value of the learning model. This BIC difference of the baseline's BIC minus the learning model's BIC can be expressed as

$$\Delta\text{BIC} = 2 \cdot [\text{LL(model)} - \text{LL(Baseline)}] - d \cdot \log(N) \tag{6}$$

where d is the difference of a learning model's number of parameters minus the statistical baseline model's number of parameters and with the Log Likelihood computed over all choices. If the learning model predicts the behavior better than the baseline model, then positive values for ΔBIC result. The larger the positive values for ΔBIC the better the learning model predicts the observed learning process. Fig. 4 shows participants' choices of the four decks and the predicted choices of the five decay-learning models.

As the first step in model comparison we tested whether the learning models were better than the statistical baseline model. We first examined the merit of the Bayesian advice integration model. The average negative values of $\Delta\text{BIC} = -8.29$ ($Mdn = -5.49$, $SD = 18.03$) of the Bayesian advice integration model shows that the model does worse in predicting choices than the statistical baseline model. Moreover, it does worse than any of the decay models. The estimated average parameter values were $\phi = 51.55$ ($Mdn = 55.64$, $SD = 5.52$) for the learning rate; $\gamma = 5.52$ ($Mdn = 6.31$, $SD = 3.41$) for the sensitivity; and $\tau = 26.27$ ($Mdn = 4.32$, $SD = 34.25$) for the higher initial evaluation of the recommended option. This result confirms previous findings that Bayesian learning models do worse in predicting experience-based decision processes in comparison to reinforcement learning models (e.g., Busemeyer & Stout, 2002; Gans et al., 2007; McElreath et al., 2005; Yechiam & Busemeyer, 2005). Due to the poor performance of the Bayesian model we restrict the remainder of the model analysis to the learning models that assume decay processes. Table 2 shows that the decay models have, on average, positive ΔBICs, indicating that they performed better than the statistical baseline model even when taking their complexity into account. We next examined whether the social learning models performed better than the individual RL model, which had a mean ΔBIC of 2.55. The mean ΔBICs for the social learning models were 7.08 for ARC-Initial, 10.77 for the outcome-bonus model, 8.13 for ARC-Decay, and
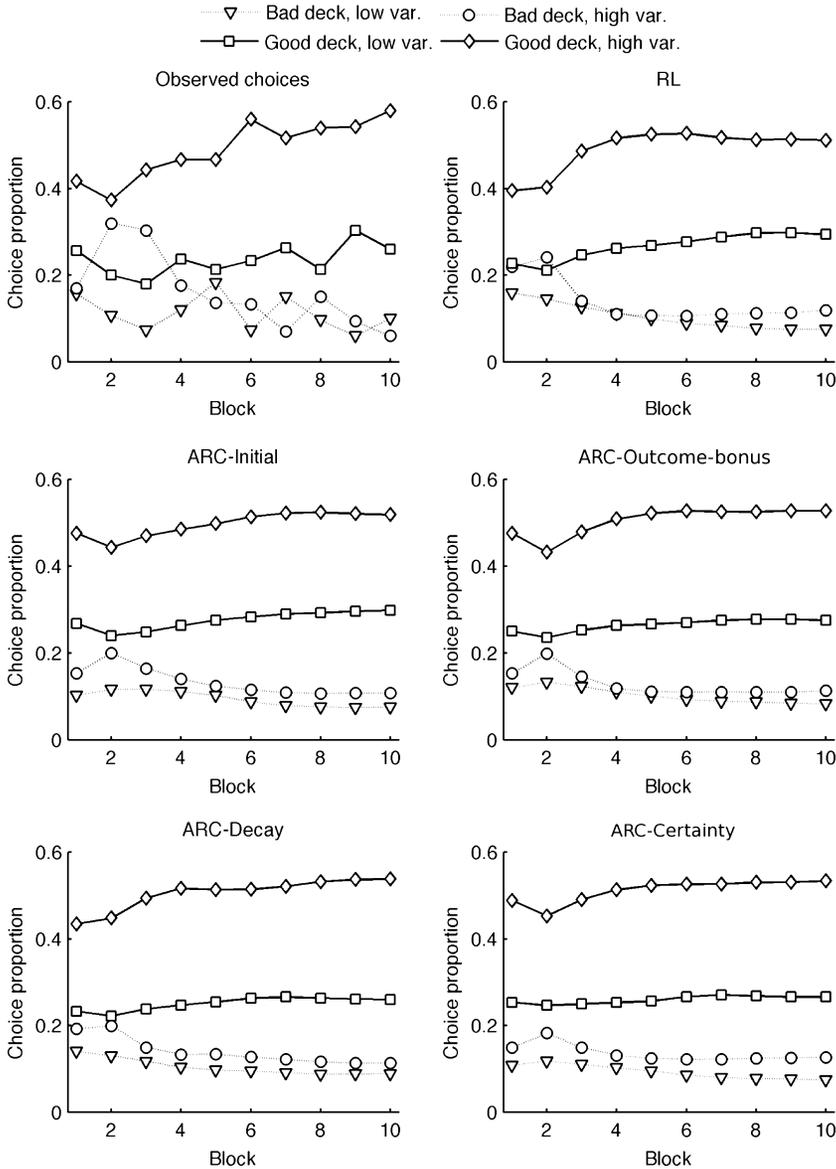
Fig. 4. Receivers' observed and predicted choice proportions for the four decks in Experiment 1 (in blocks of 10 trials). The first panel shows observed choice proportions. The other panels show the models' predictions from the simulation. Each participant was simulated with the best parameter set before individual level predictions were averaged (see text for details). The legend at the top is valid for each of the six panels. RL, reinforcement learning model.

10.4 for ARC-Certainty (see Table 2 for details). The *t*-tests (Table 3) illustrate that the social learning models describe the learning process better than the individual RL model. Together with the previous finding that the recommended deck was chosen more frequently

Table 2
Mean (median, *SD*) for ΔBIC and parameter values for the individual reinforcement learning (RL) model and the advice-reinforcement combination (ARC) models

Model Fit or Parameter	RL	ARC Models			
		Initial	Outcome-Bonus	Decay	Certainty
ΔBIC	2.55 (1.59, 23.96)	7.08 (4.02, 21.69)	10.77 (3.99, 16.56)	8.13 (4.87, 18.66)	10.4 (7.06, 15.26)
Social learning parameter		ι = 21.18 (12.05, 23.22)	ρ = 3.89 (2.95, 3.88)	$\phi_{advised}$ = .17 (.08, .29)	τ = .16 (.13, .14)
Decay (ϕ)	.16 (.05, .26)	.18 (.05, .3)	.36 (.2, .32)	.52 (.57, .35)	.32 (.1, .36)
Sensitivity (λ)	3.26 (4.42, 2.09)	1.55 (1.63, 1.59)	3.1 (4, 2.05)	3.09 (3.93, 2.02)	2.44 (2.82, 2.3)
RMSD					
Good deck	.062	.055	.059	.056	.063
Adherence	.102	.048	.036	.036	.040
All choices	.048	.044	.045	.043	.046

Note: Social learning parameters are additional initial attraction (ι) for ARC-Initial, additional reinforcement (ρ) for ARC-Outcome-bonus, separate decay ($\phi_{advised}$) for ARC-Decay, and standard deviation threshold (τ) for ARC-Certainty. *RMSD* is the mean over all trials of the squared deviation between predicted and observed choice proportions on the group level. For ''Good deck,'' all choices of a good deck were considered; for ''Adherence,'' all choices in which the recommended deck was chosen were considered; and for ''All choices,'' all choices were considered.

Table 3
Comparisons of the individual reinforcement learning (RL) model and the advice-reinforcement combination (ARC) models according to their ΔBIC values by mean of *t*-tests

Models	ARC-Initial	ARC-Outcome-Bonus	ARC-Decay	ARC-Certainty
RL	*t* = 1.95, *p* = .060, *d* = .2	*t* = 3.34, *p* = .002, *d* = .39	*t* = 2.68, *p* = .012, *d* = .26	*t* = 2.71, *p* = .011, *d* = .38
ARC-Initial		*t* = 2.03, *p* = .051, *d* = .19	*t* = .46, *p* = .648, *d* = .05	*t* = 1.65, *p* = .110, *d* = .18
ARC-Outcome-bonus			*t* = −2.21, *p* = .035, *d* = −.15	*t* = −.39, *p* = .700, *d* = −.02
ARC-Decay				*t* = 1.38, *p* = .178, *d* = .13

Note: For all tests *df* = 29. When *t* statistics are negative (positive), the row model is better (worse) than the column model.

than the corresponding deck, this result supports the assumption that a social learning process describes decision makers' choices better than purely individual learning.

To examine whether one social learning model outperformed the other ARC models, we conducted *t*-tests comparing the models' ΔBICs (see Table 3). A comparison of the social learning models shows that the outcome-bonus model has the best fit in describing the observed learning process in comparison to all other models. However, the fit differences are small: ARC-Outcome-bonus is not significantly better than ARC-Certainty, and the effect sizes of the comparisons with ARC-Initial and ARC-Decay are small ($d = .19$ and $d = .15$, respectively). In sum, the social learning models explain participants' choices better than a statistical baseline model and, importantly, better than the individual RL model. Among the social learning models, the outcome-bonus model predicted the observed learning process best. However, the small effect size of the differences in fit between ARC-Outcome-bonus and the competing models highlights that ARC-Decay and ARC-Certainty also did quite well. This is also reflected in the root mean square deviation (*RMSD*) as an alternative goodness-of-fit measurement (see bottom of Table 2).

Considering the social learning parameters, the median ι for ARC-Initial was 12.05—that is, according to this model the initial attraction for the recommended decks was approximately 12 times the average payoff from a good deck (.125 cents); the median ρ for ARC-Outcome-bonus was 2.95—that is, every reinforcement from a recommended deck received a ''bonus'' that was equivalent to three times the average payoff from a good deck; and the median ϕ_{advised} for ARC-Decay was .08, clearly lower than the decay rate of .57 for options that were not recommended. In ARC-Certainty the median of the social learning parameter τ was .13. That is, participants chose the recommended option, on average, when the standard deviation of the choice probabilities was below $\tau = .13$.

The social learning models differ from the individual learning model by assuming additional social learning mechanisms. However, we did not only estimate the parameter representing the social mechanism but estimated all parameters for the social learning model to give the models their full power to describe the observed learning processes. A potential drawback of this method is that the social learning models potentially were able to predict the learning process better due to a *combination* of social and standard learning mechanisms. Therefore, we examined whether the estimated parameters of the individual learning model that are also part of the social learning model differ for the two type of models. The average parameter values for the individual learning model fitted to the independent decision makers were $\gamma = 2.69$ (*Mdn* = 2.32, *SD* = 2.12) for sensitivity and $\phi = .29$ (*Mdn* = .14, *SD* = .33) for the decay rate. Table 4 shows the results of comparing the standard learning parameters for the independent decision makers with the standard learning parameters for receivers. It becomes clear that the parameters estimated do not differ substantially. The exceptions are the sensitivity parameter for ARC-Initial and the learning rate for ARC-Decay.

We also examined individual differences in how people use advice. We found that five participants made all of their 100 choices consistent with the advice. We also explored for how many participants a specific model was the best. The outcome-bonus model best describes most participants (10), followed by ARC-Certainty (seven), RL (six), ARC-Decay (four), and ARC-Initial (three).

Table 4
Mann–Whitney *U*-tests for comparing the estimated parameter values for the independent decision makers with parameters estimated for the advice receivers

| Parameter | RL | ARC Models | | | |
		Initial	Reinforcement	Decay	Certainty
Decay	$U = 347$, $p = .128$	$U = 343$, $p = .114$	$U = 376$, $p = .274$	$U = 274$, $p = .009$	$U = 429$, $p = .756$
Sensitivity	$U = 346.5$, $p = .126$	$U = 332.5$, $p = .082$	$U = 342.5$, $p = .111$	$U = 367.5$, $p = .222$	$U = 414.5$, $p = .599$

Note: Each cell depicts the test statistics for a comparison of the best-fitting model parameters for advice receivers for the individual (RL) and the social learning (ARC) models with the best-fitting parameters of the individual learning model applied for the independent decision makers. Because the distribution of parameter values was frequently not normal, Mann–Whitney *U*-tests were applied.
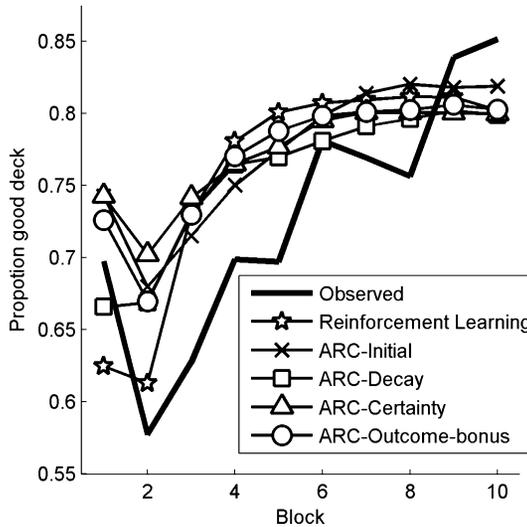

Fig. 5. Receivers' observed and predicted choice proportions for choosing one of the good decks in Experiment 1 (in blocks of 10 trials). The figure shows that the outcome-bonus model and ARC-certainty are best able to capture the nonmonotonic trend of choice probabilities.

Apart from comparing the model fits, one can query whether the models can predict characteristic patterns of choices over time. Figs. 2 and 3 show that receivers showed an adherence–exploration–adherence choice pattern. ARC model predictions were calculated by first simulating each participant 100 times with the best parameters for this participant and then averaging the resulting choice probabilities over all 30 participants.[5] Fig. 5 compares the probabilities of simulated and real participants choosing one of the two good decks in the IGT. To evaluate the correspondence of simulated and observed choices, we calculated the *RMSD* (see Table 2) between predicted and observed average probabilities on the group level. The *RMSD*s for "good decks" indicate that all social learning models predict the
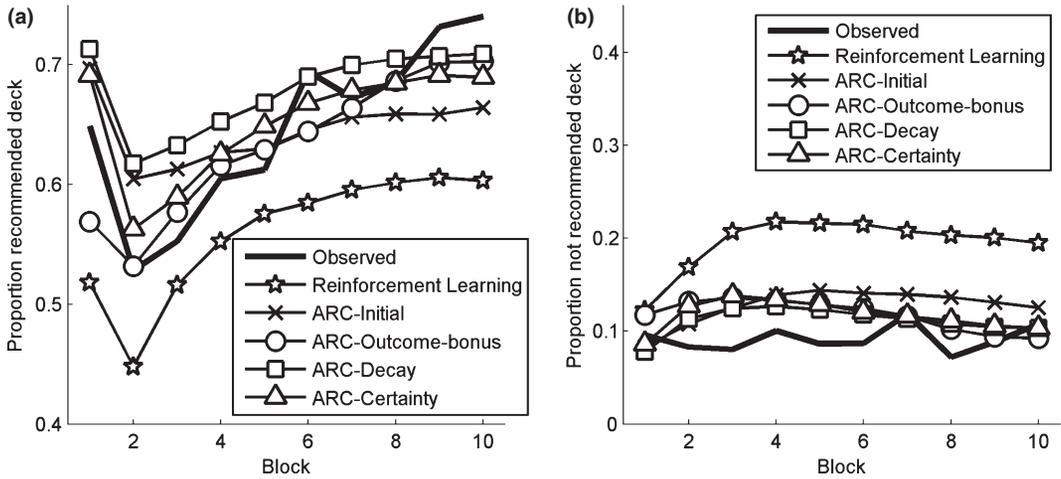
Fig. 6. Receivers' observed and predicted choice proportions in Experiment 1 for the recommended deck (a) or the corresponding deck with the same average payoff (b). The individual learning model does not capture the influence of advice, whereas the ARC-outcome-bonus model and the ARC-decay model do so well.

proportion of choices of good decks similarly well (see also Fig. 5). However, this result changes when differentiating the recommended from the nonrecommended option(s). Fig. 6 shows the probabilities of simulated and real participants choosing a good or bad deck, given whether it was recommended or not. For instance, when a participant was advised to choose the good deck C, we calculated the probability that this deck would be chosen and the probability that the corresponding equally good deck D would be chosen. Fig. 6 and the *RMSD*s in Table 2 show that the ARC models implementing social learning—especially ARC-Outcome-bonus, ARC-Decay, and ARC-Certainty—can better account for the adherence–exploration–adherence pattern because they are better able to describe the rebound of choices of recommended options after the exploration phase.

5.3. Discussion of Experiment 1

Experiment 1 illustrated that people use advice to improve their decisions. However, only a few receivers (i.e., five) followed the advice for all 100 choices. The majority of receivers chose the recommended option first but then explored the other options, frequently returning to the recommended option later in the experiment. Still, receivers performed better than independent decision makers, who made decisions without receiving or giving advice, and better than participants who had to give advice (advisors). Compared to these groups, receivers were approximately 10% points more likely to choose a good deck. However, receivers (73% chose a good deck) performed worse than participants with their own experience in the same task (78% chose a good deck). In sum, receivers had an advantage over inexperienced decision makers but did worse than experienced individuals (i.e., advisors in their last 100 trials). Surprisingly, advisors did not start with the same deck they recommended to others when they made decisions in their second 100

trials of the IGT. It could be that due to the break between the two parts of the experiments advisors expected the environment to change, even though they were told by the experimenters that they would be choosing from the same card decks again. Alternatively, the participants might have learned that the outcomes of the first several cards of the two bad decks were not that bad.

The model comparisons showed that the Bayesian advice integration model could not describe choices well, perhaps because the Bayesian model cannot account for the strong influence of recent outcomes, which is reflected in the high decay parameters estimated for the decay models (see Table 2). Examining ΔBICs, we found that the other social learning models performed better than the pure individual learning model. Of the social learning models, we found that ARC-Initial was, on average, least able to describe participants' choices. This is reflected in the worst fit and more importantly, in the weaker ability of ARC-Initial to account for the characteristic adherence–exploration–adherence choice pattern of many participants. It is less clear which of the three remaining social learning models did best in predicting the learning process. Even though ARC-Outcome-bonus had a better fit than ARC-Decay and ARC-Certainty, the difference was small compared to the ARC-Initial model. Therefore, we designed Experiment 2 to provide a more specific test of the best models that emerged in Experiment 1. A peculiarity of Experiment 1 was that participants rarely received bad advice; therefore, in Experiment 2 we additionally examined how bad advice influences learning, and whether the social learning models can describe learning after bad advice.

6. Experiment 2

Experiment 2 focused on a comparison of ARC-Outcome-bonus, ARC-Decay, and ARC-Certainty by testing their predictions with a strong generalization test (Busemeyer & Wang, 2000). For this test we used the estimated parameter values from Experiment 1 to determine the models' predictions for the new learning situation in Experiment 2.

For a test to be strong, it is desirable to find a situation in which the three models make different qualitative predictions. Such a situation occurs for a task in which all decks have negative expected payoffs. Specifically, the outcome-bonus model predicts that participants will still prefer the recommended deck. In contrast, ARC-Decay predicts that individuals should avoid the recommended option because of the lower decay rate for a recommended deck. The decay model invokes slower forgetting of negative payoffs, which means that propensities remain negative for a longer time after losses. Compared to the recommended option, the other options will appear more attractive because the higher decay rate implies a short memory for the negative payoffs.

ARC-Outcome-bonus and ARC-Certainty both predict that receivers will prefer the recommended deck over the corresponding decks. However, only ARC-Certainty explicitly predicts stronger adherence to advice when receivers are more uncertain about which options are better. Uncertainty can be expressed by the effect size of the payoff difference of the good and bad decks. The effect size of the original IGT is $d = .24$ (see Cohen, 1988),

calculated as the difference in payoffs from good and bad decks divided by the standard deviation of the pooled payoffs from all options.

In sum, to create a situation in which the three social learning models make qualitatively different predictions in Experiment 2, a task was required in which all four options would predominantly lead to negative payoffs, allowing a distinction between the decay model and the outcome-bonus model. Additionally, the payoff difference between good and bad decks should have a small effect size, allowing a distinction between ARC-Certainty and ARC-Outcome-bonus. In line with these demands we devised a payoff schedule with negative expected payoffs and an effect size of $d = .15$ for the payoff difference between good and bad decks.

To determine the models' quantitative predictions before running the experiment, we applied a nonparametric bootstrapping procedure. Specifically, the choices of a virtual participant in Experiment 2 were simulated by using the parameters from one randomly selected real participant from Experiment 1.[6] With these parameter values learning and choices over the 100 trials was determined 100 times, and the simulated participant's expected choice probabilities for each trial were determined as the mean over the 100 simulations. To obtain average choice probabilities for one ''virtual experiment,'' the parameters of 30 randomly selected participants (with replacement) from Experiment 1 were matched with random advice, so that on average 50% of the advice was good. The models' predictions in Experiment 2 were determined by averaging over 5,000 virtual experiments. Fig. 7 displays the three models' predictions for adherence to advice in Experiment 2.
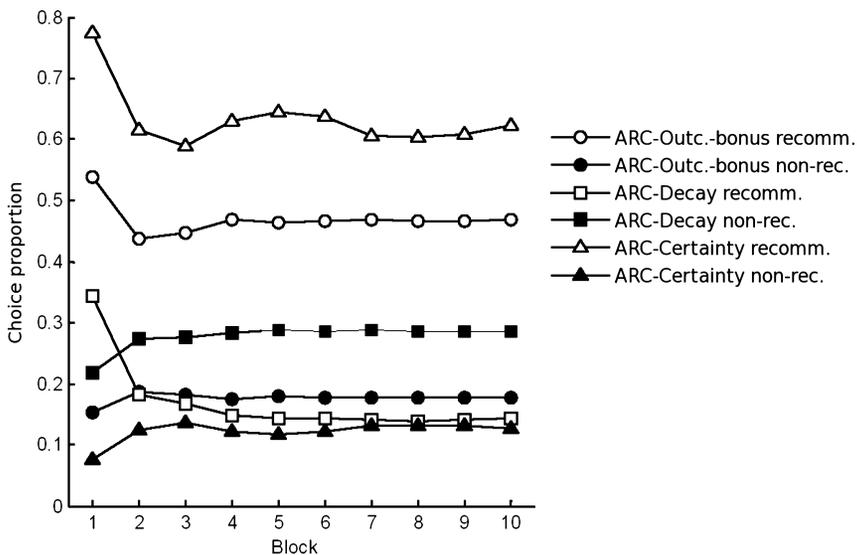

Fig. 7. Predicted choice proportions for Experiment 2 according to different social learning models average for 5,000 simulated experiments. For each simulated experiment the parameter values of 30 randomly drawn participants with replacement from Experiment 1 were used to determine the models' predictions (see text for details). The models make different quantitative and qualitative predictions: ARC-certainty predicts high adherence rates, the outcome-bonus model predicts lower adherence rates, and only the decay model predicts that the recommended deck will be chosen less frequently.

6.1. Design

Experiment 2 used a learning task similar to that used in Experiment 1. The key difference was the payoff schedule: average payoffs in Experiment 2 were –10 eurocents for the bad decks and –7 eurocents for the good decks, with a standard deviation of 20 eurocents for all decks. Payoffs were randomly drawn from a discrete approximation of a normal distribution, and mean and standard deviation of payoffs were maintained in blocks of 15 trials; we call this the standard payoff schedule. Participants started with an initial endowment of 12.5 euros and made 105 choices.

One result from Experiment 1 was that participants rarely received bad advice. Pilot tests for Experiment 2 revealed that it was difficult to find a payoff schedule under which approximately half of the participants would learn which were the better decks and, hence, give good advice. Therefore, we made the task for 20 of the 30 advisors in Experiment 2 more difficult by manipulating their payoff schedule (henceforth, nonstandard advisors) so that approximately half of the participants would receive good advice and the other half bad advice. The manipulation consisted of subtracting 5 eurocents from every payoff from a good deck in the first 30 trials, and adding 5 eurocents to the same deck in 30 randomly selected trials from the last 75 trials.[7] As in Experiment 1, the task was performed by advisors, receivers, and independents, with the latter two always choosing from the standard payoff schedule.

6.2. Participants and procedure

Eighty participants, mostly students from the Free University of Berlin (55% women, mean age 25 years), were randomly assigned to the three conditions. Thirty participants were advisors, 30 were receivers, and 20 were independents. As in Experiment 1 participants received information only about the stochastic nature of the payoff distribution, and no information about the domain (gain or loss) or variability of the payoffs. With two exceptions, the experimental procedure was identical to Experiment 1. First, the similarity of decks for advisors and receivers was expressed as decks having the same average payoff, instead of describing them as being identical, because in the case of the advisors, the good decks started with lower expected outcomes and thereafter improved. Second, participants' variable payoff was calculated without subtracting the initial endowment because the expected payoff for all decks was negative.

6.3. Results

6.3.1. Choices and performance

Participants earned, on average, 3.83 euros ($SD = 1.9$). Fig. 8 shows the proportion of participants who chose one of the two good decks and shows that the task in Experiment 2 was more difficult for most participants because the average proportion of participants choosing a good deck was, compared to Experiment 1, lower throughout the task. The 10 standard advisors chose a good deck, on average, in 60% ($SD = 17\%$) of their first 100 trials and in 69% ($SD = 10\%$) of their second 100 trials. The 20 nonstandard advisors chose a
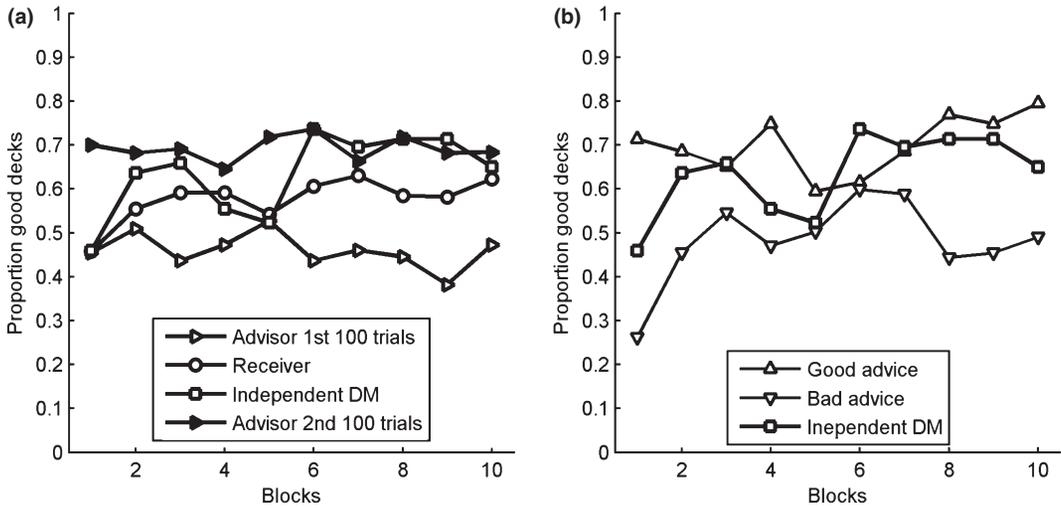
Fig. 8. Participants' average choice proportions of choosing one of the good decks in Experiment 2. (a) Comparison of independents, advisors, and receivers. (b) Comparison of independents and receivers of good or bad advice.

good deck, on average, in 39% ($SD = 12\%$) of their first 100 trials and in 42% ($SD = 40\%$) of their second 100 trials. This poorer performance of nonstandard advisors was expected because their task was more difficult. Of the receivers, 13 received the good advice to choose from the good decks and the remaining 17 received bad advice. The 13 receivers of good advice chose a good deck, on average, in 69% ($SD = 15\%$) of their trials, the 17 receivers of bad advice chose a good deck 48% ($SD = 20\%$) of the time, and the independents chose a good deck in 63% ($SD = 12\%$) of their trials. Comparing advisors, receivers, and independents, we found that receivers of good advice performed better than those who received bad advice, $t(28) = 3.08$, $p = .005$, $d = 1.1$, but not better than independents, $t(31) = 1.25$, $p = .22$, $d = .45$, or advisors in their second 100 trials, $t(21) = .03$, $p = .97$, $d = .01$. Receivers of bad advice performed worse than independent decision makers, $t(35) = 2.73$, $p = .01$, $d = .9$.

Over all trials, independent participants chose one of the good decks, on average, in 67% ($SD = 13\%$) of the trials, which is significantly higher than 50%, $t(19) = 4.71$, $p < .001$, $d = 1.05$, and suggests that they learned which decks produced lower losses. They also chose one of the good decks in 70% ($SD = 9\%$) of the last 50 trials, which suggests the same conclusion. In sum, participants' choices show that the manipulation successfully made the task more difficult for nonstandard advisors. In the more difficult task in Experiment 2, the receivers benefited less from the advice, due to frequent bad advice. Independents still learned to choose the good decks.

6.3.2. Giving and following advice

Three of 10 standard advisors and 14 of 20 nonstandard advisors (with a more difficult task) recommended choosing a bad deck. Most advisors recommended the deck they had
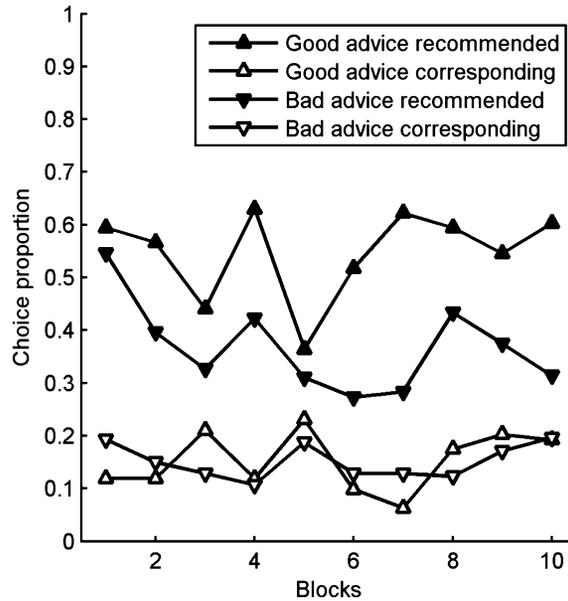
Fig. 9. Proportion of participants choosing a deck conditional on advice in Experiment 2 (blocks of 10 trials). Even for participants who received bad advice a strong influence of advice on choice was observed.

chosen most frequently (on average in 35%, $SD = 14\%$, of the trials), indicating that they recommended the deck they preferred themselves. Fig. 9 displays the proportion of receivers who chose the recommended or the corresponding deck (see also Table 5). To test the influence of the quality of advice, we performed a repeated measures analysis of variance, with the quality of advice (good vs. bad) as a between-subjects factor, advice (recommended deck vs. corresponding deck) as a within-subject factor, and the choice frequency of recommended and corresponding decks as a dependent variable. Participants who received good advice chose the recommended deck in on average 55% of all trials ($SD = .19$) and the corresponding deck in only 15% of all trials ($SD = .13$). When receiving bad advice the corresponding choice proportions were 37% ($SD = .19$) for choosing the recommended option and 15% ($SD = .08$) for choosing the corresponding deck.

The statistically significant main effect for advice, $F(1,28) = 47.51$, $p < .001$, $\eta^2 = .63$, shows that the recommended deck was chosen more frequently than the corresponding deck. The statistically significant main effect for quality of advice, $F(1,28) = 6.38$, $p = .017$, $\eta^2 = .19$, shows that receivers adhered more to advice when it was good. The interaction between quality of advice and advice was not statistically significant, $F(1,28) = 3.78$, $p = .063$, $\eta^2 = .12$, and the small effect size suggests that the probability of choosing the corresponding deck was not substantially influenced by the quality of advice. The larger effect size for advice in comparison to the effect of the quality of advice, indicates that participants' choices were more influenced by advice than by the payoffs from choosing the decks.

Table 5
The predicted choice probabilities of choosing a good, recommended, and corresponding deck according to the bootstrapping simulation. The predicted choice probabilities are compared to the actual observed choice proportions through the root mean square deviation (*RMSD*)

Advice	Deck	Observed Choices	ARC				
			RL	Initial	Outcome-Bonus	Decay	Certainty
Good	Good	.7 (.69, .15)	.57 (.57, .04)	.61 (.59, .07)	.67 (.67, .16)	.46 (.45, .05)	.79 (.78, .19)
	Recommended	.55 (.53, .18)	.29 (.28, .02)	.36 (.34, .11)	.49 (.48, .25)	.18 (.15, .08)	.65 (.64, .31)
	Corresponding	.15 (.12, .13)	.28 (.28, .02)	.25 (.26, .04)	.18 (.18, .09)	.28 (.28, .03)	.14 (.14, .13)
	RMSD						
	Good deck	–	.138	.102	.061	.238	.092
	Adherence	–	.207	.157	.073	.279	.082
	All choices	–	.112	.089	.069	.149	.077
Bad	Good	.48 (.48, .2)	.55 (.57, .05)	.51 (.54, .06)	.38 (.39, .18)	.57 (.58, .07)	.28 (.29, .26)
	Recommended	.37 (.37, .19)	.23 (.22, .03)	.29 (.25, .08)	.45 (.44, .25)	.16 (.14, .11)	.62 (.6, .35)
	Corresponding	.15 (.17, .08)	.22 (.21, .03)	.2 (.2, .03)	.17 (.17, .08)	.27 (.28, .03)	.11 (.1, .09)
	RMSD						
	Good deck	–	.101	.058	.104	.091	.202
	Adherence	–	.12	.081	.067	.177	.177
	All choices	–	.105	.079	.084	.112	.127
Good and bad	Good	.58 (.55, .21)	.56 (.57, .05)	.55 (.56, .08)	.51 (.51, .22)	.52 (.53, .08)	.5 (.51, .34)
	Recommended	.45 (.39, .2)	.26 (.24, .04)	.32 (.29, .1)	.47 (.46, .25)	.17 (.14, .1)	.63 (.62, .34)
	Corresponding	.15 (.13, .1)	.24 (.25, .04)	.22 (.24, .04)	.17 (.18, .08)	.27 (.28, .03)	.12 (.12, .11)
	RMSD						
	Good deck	–	.046	.031	.068	.057	.074
	Adherence	–	.155	.111	.048	.22	.133
	All choices	–	.085	.068	.067	.089	.068

Note: Cells depict means (medians, standard deviations); predicted standard deviations were calculated as the mean of the standard deviations over all simulated experiments.

6.3.3. Model comparison

According to the outcome-bonus model, participants who received bad advice should choose a bad deck more often than participants who received no advice. This can also be assumed for the certainty model, because the high variance of payoffs from decks was high, so that participants should rely on advice. In contrast, the decay model predicts that receiving bad advice should decrease the probability that receivers will choose the recommended bad deck. As Fig. 8 depicts, unlike the prediction from the decay model, participants who received bad advice chose the good decks less frequently than independent decision makers. Further, as predicted by the outcome-bonus model and ARC-Certainty, receivers chose a nonrecommended deck less frequently than the recommended deck, contrary to the decay model's prediction. Finally, the proportions of choices of the recommended decks did not increase substantially from Experiment 1 to Experiment 2, as predicted by the certainty model. Hence, the examination of the qualitative predictions speaks in favor of the outcome-bonus model.

Additionally, to compare the models quantitatively, we examined how well they predicted participants' choices. The models' predictions were simulated as described above, except that parameter values from participants in Experiment 1 were randomly matched with real recommendations from Experiment 2. To compare the models, we examined how well the models predicted the choices of the recommended deck and the corresponding deck with the same expected payoff. Fig. 10 shows the observed and simulated choice proportions and illustrates that the outcome-bonus model ($RMSD$ = .048) predicted the preference for the recommended decks better than ARC-Decay ($RMSD$ = .22) and also better than ARC-Certainty ($RMSD$ = .133). ARC-Certainty overestimated the influence of social
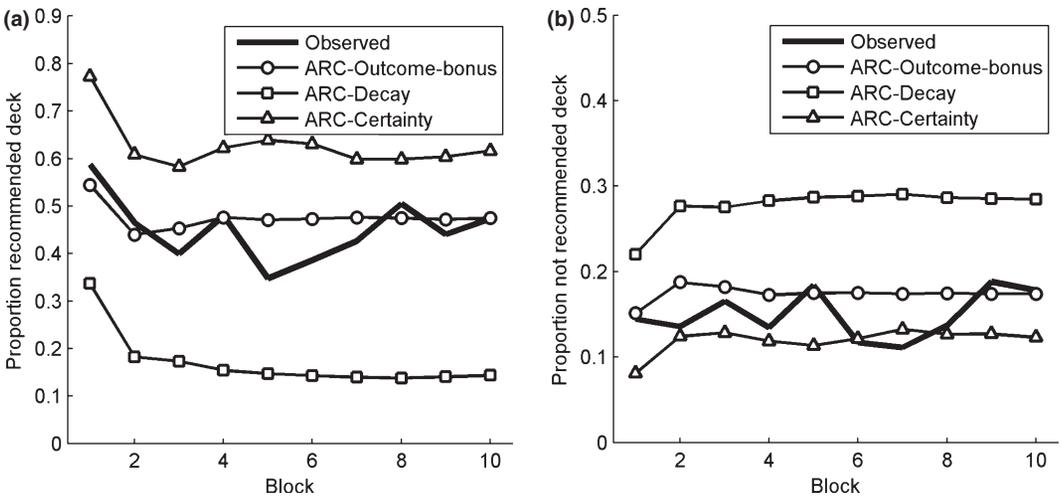

Fig. 10. Receivers' observed and predicted choice proportions in Experiment 2 for (a) the recommended deck or (b) the corresponding deck with the same average payoff. Only the outcome-bonus model predicts choices well, whereas the decay model overestimates the preference for the not-recommended deck and ARC-Certainty overestimates the preference for the recommended deck.

learning in Experiment 2. Table 5 shows that ARC-Outcome-bonus predicted mean choice proportions better than ARC-Decay and ARC-Certainty and also predicted the standard deviations of choice proportions within a group observed in Experiment 2. Thus, the generalization test supports the outcome-bonus model as the best social learning model considered.

To investigate whether good advice is treated differently from bad advice, we performed the same bootstrapping procedure as above, separately for receivers of good and bad advice. Table 5 shows the results of these simulations, which demonstrate that with good advice the outcome-bonus model predicted adherence to advice well, whereas the decay model underestimated adherence as well as the proportion of choices of good decks, and ARC-Certainty overestimated both. With bad advice ARC-Decay again underestimated adherence to advice and again chose—as predicted—a nonrecommended deck more frequently than the recommended deck. ARC-Certainty clearly overestimated adherence to advice and, hence, predicted low proportions of choices of the good decks. To a lesser degree this was also true for the outcome-bonus model, which, nevertheless, still correctly predicted stronger adherence to good than to bad advice, even though the predicted difference of 4% is smaller than the observed difference of 18%. In sum, the comparison of the models' predictions for good and bad advice also shows that the outcome-bonus model is the best social learning model among the set of tested models.

6.3.4. Discussion of Experiment 2

Experiment 2 allowed us to examine social learning in a situation in which the best social learning models from Experiment 1 made different predictions. We found again that receivers generally used advice. Supporting the results in Experiment 1, we found that advice had a greater impact on receivers' choices compared to the payoffs they received from their choices. Furthermore, good advice improved performance and bad advice harmed performance.

Experiment 1 showed that advisors did not choose the decks that they had recommended when they started with their second 100 trials. In contrast, advisors in Experiment 2 did do so. This suggests that participants trusted the instructions, and that advisors in Experiment 1 had learned that one could choose from the bad decks in the early trials without risking high losses.

Experiment 2 rejected the qualitative prediction of the decay model: participants receiving bad advice chose bad decks more frequently than independent participants. Further, receivers chose the recommended deck more frequently than the corresponding deck with the same expected payoff. These results are in line with the predictions from the outcome-bonus model and ARC-Certainty. The simulated predictions of the models in Experiment 2 favored the outcome-bonus model because ARC-Certainty predicted a too high adherence to advice. Only the outcome-bonus model correctly predicted receivers' adherence to advice and predicted more adherence to good than to bad advice, even though the predicted difference is smaller than the observed difference. Finally, the variance in participants' probability of choosing a good deck simulated with the outcome-bonus model using parameter estimates from Experiment 1 is similar to the observed variance in Experiment 2.

Experiment 1 showed that different participants are best modeled by different learning models. As the models were not fitted to participants in Experiment 2, it is not possible to classify participants by model. We nevertheless suggest that of the models considered, the outcome-bonus model is not only best on average but also best describes most participants. This argument is consistent with the finding that participants who received bad advice did not chose the bad decks less frequently at the end of the experiment; instead the choice proportion of the bad decks stayed constant for the last 70 trials. In contrast to this empirical observation, ARC-Initial and ARC-Decay predict a decreasing choice probability. Interestingly, in contrast to Experiment 1, no participant in Experiment 2 followed the advice in all 100 trials. While this could be a coincidence, participants in Experiment 2 experienced losses in most trials, and this might have stimulated a stronger exploration of alternative options. In sum, Experiment 2 clearly supports the outcome-bonus model as the best of the models considered to describe the social learning process because it predicted adherence to advice, conditional on the quality of advice, and also predicted the variance of choice proportions.

7. General discussion

We examined social learning in the context of repeated choices from experience. We aimed to answer three questions: Do people use advice? Does taking advice improve decision performance? How can social learning be best described? To answer these questions, we observed choices in tasks with repeated choice among four options and tested one model of individual and five models of social learning.

First, we found that receivers used advice as evidenced by the fact that they chose the recommended deck more frequently than the corresponding deck with the same average payoff. Moreover, in Experiment 2, receivers even followed the bad advice to choose decks with the lowest expected payoff. The influence of advice was also visible in task performance: as in Experiment 1, receivers of good advice performed better than independent decision makers, who were, in turn, better than participants receiving bad advice. The poor performance of receivers of bad advice in Experiment 2 shows that social influence can distract people from solving a task on the basis of individual learning. However, participants adhered more to good advice than to bad advice, showing that individual learning still played a role. Overall, this suggests that individual experience was combined with the social information to inform choices.

7.1. Social learning models

We proposed and tested five models of social learning and compared them with an individual learning model. The individual learning model represents a simple reinforcement model (e.g., Erev, 1998; Yechiam & Busemeyer, 2005). Four social learning models are modifications of the individual learning model and one social learning model represents a Bayesian approach. ARC-Initial assumes that the recommended options are

initially evaluated more positively, compared to alternative options. ARC-Outcome-bonus assumes that payoffs from recommended options lead to stronger reinforcement. The decay model assumes that propensities of recommended options decay more slowly. ARC-Certainty assumes that people choose the recommended options when the propensities of options are similar. Finally, the Bayesian model assumes that the recommended choice option has a higher prior expected reward and uses Bayes' rule to integrate new information.

Experiment 1 showed that three of the five social learning models—ARC-Outcome-bonus, ARC-Decay, and ARC-Certainty—described choices better than the statistical baseline model and the individual reinforcement learning model. The Bayesian model did worse than the statistical baseline model and all other social learning models, so it is disqualified as an appropriate description of the observed learning process.

In Experiment 2 ARC-Outcome-bonus, ARC-Decay, and ARC-Certainty were further tested against each other using a modified version of the IGT with negative average payoffs and high payoff variance between and within choice options. In this situation, the three best models identified in Experiment 1 made diverging predictions. In accord only with the predictions of ARC-Outcome-bonus and ARC-Certainty, participants consistently chose the recommended option, which in the case of bad advice means that they did not find the best option. Whereas ARC-Certainty generally overestimated adherence to advice, the outcome-bonus model correctly predicted that adherence to advice is higher when advice is good and also predicted the variance of choice proportions in Experiment 2. In sum, the experiments show that decision makers adhere to advice, that good advice helps and bad advice harms learning, and that the outcome-bonus model provides the best description of the social learning process.

When social learning diverges from individual learning, social information ought to influence individual learning in a different way than one's own experience. The finding that ARC-Initial and the Bayesian model did not adequately model the learning process indicates that accurate models do not assume that advice only influences the decision makers' initial preference, as assumed by models of individual learning that account for decision makers' prior experience (Camerer et al., 2002; Hanaki et al., 2005). The conclusion that advice influences learning differently from one's own experience is also supported by the finding that advisors in their second 100 trials behaved differently from receivers. As Figs. 2 and 8 show, receivers explored alternative options longer than advisors did in their second 100 choices.

How does social learning differ from individual learning? The outcome-bonus model provided the best account of the observed learning process. According to this model, social learning differs from individual learning by evaluating the outcome of recommended options more positively in comparison to the outcome of nonrecommended options. Thus, the outcome-bonus model predicts that individuals subjectively experience an outcome of a recommended option more positively than the identical outcome of a nonrecommended option. Because the advice bonus accumulates in the options' propensities, the impact of social information on the choice probabilities does not diminish but rather increases over time. Importantly, a bad recommendation, due to its influence on the learning process, leads

to poorer performance than an individual learning process under no social influence. In sum, social information determines choices briefly at the beginning of a task, and more strongly and persistently after the exploration phase.

As the results in Experiment 2 show, participants' dependence on the quality of advice has the disadvantage of impairing performance when advice is bad. Hence, decision makers should possess mechanisms to attenuate the effect of bad advice; for instance, they should be very selective when choosing advisors. Whereas our study was arranged so that participants could assume that advisors were competent, the experiments of Yaniv and Kleinberger (2000) and Luan et al. (2004) showed that, if possible and necessary, receivers reacted sensitively to the quality of advice. They disregarded advice from advisors who repeatedly gave bad advice, thus hedging the sensitivity of their social learning mechanism to bad advice. Experiments by Celen, Kariv, and Schotter (2005), Kameda and Nakanishi (2003), and Yaniv (2004a) showed that even naïve participants tend to give useful advice, and that social learning generally improves performance.

7.2. Limitations of the learning models

Our models describe the social learning process at the computational level, but some questions remain open due to limitations in our experiments. One unresolved issue is whether all decision makers can be described with the same model or whether people have qualitatively different learning processes. Results in Experiment 1 suggest that different participants are best described with different models, whereas the analysis of Experiment 2 suggests the ARC-Outcome-bonus model predicted the behavior of most receivers well. The good performance of ARC-Outcome-bonus in predicting mean and variance of choice probabilities in Experiment 2 suggests that this model is sufficient to predict most individuals' behavior. To explore whether stable individual heterogeneity in learning processes exists, for which different learning models are required, it will be necessary to examine participants' behavior when performing the same basic task repeatedly.

Experiments in which participants perform tasks with different payoff distributions could also help remedy a second limitation, namely, that our selection of ARC-Outcome-bonus is based on two specific payoff distributions. These distributions vary important payoff characteristics (positive and negative expected values and different levels of payoff variance), but many other distributions exist and for some of them ARC-Outcome-bonus makes counterintuitive predictions. For instance, if the payoff for option A is always 20 and for option B in 99% of the cases is 21 and in 1% of the cases −1,000, then an advisor with enough experience would recommend choosing option A. Here ARC-Outcome-bonus makes the counterintuitive prediction that people would converge to choose option A after a short exploration phase, whereas ARC-Certainty would predict that choices converge on option B (until the first negative payoff is experienced). These counterintuitive predictions are worth studying in order to generalize the reported findings.

A third limitation concerns ARC-Outcome-bonus's overestimation of adherence to advice after receiving bad advice. While this suggests that the individual learning component of participants' behavior in Experiment 2 was underestimated, this is probably not due

to the general inability of ARC-Outcome-bonus to explain their behavior. Rather, underestimation occurred because nearly all participants in Experiment 1 received good advice, and so individual and social learning usually pointed in the same direction, allowing for relatively high values for the social learning parameters. The high social learning parameter values might then have suppressed ARC-Outcome-bonus's individual learning component in the simulations used to predict behavior in Experiment 2.

A final limitation concerns all the learning models tested. As Figs. 5 and 6 illustrate, the models predicted (to different degrees) the general trend of participants' choices but were less able to account for local fluctuations in choice probabilities that also characterized the learning process. The smoothness of the models' predicted curves results, in part, from averaging across 5,000 simulations. Beyond that, the higher variance in the observed data might indicate that learning includes more than simple reinforcement processes.

7.3. Generalization to other learning situations

The generalizability of these results depends partly on how advice is given in other contexts and the incentives involved. In the experiments presented here, receivers were advised to choose a particular option and were aware that advisors were paid according to the receivers' performance. One might question whether advice that is less strict might lead to less adherence to advice. Using a similar task, Biele, Rieskamp, Krugel, and Heekeren (unpublished data) found that when participants were advised to ''mostly choose'' one particular option, adherence to advice was still generally high even though fewer receivers exclusively chose the recommended option. Furthermore, in our experiments the receivers were aware that the advisors benefit from receivers' decisions. We used the advisors' performance-dependent payment to signal the advisors' motivation to receivers, but we cannot exclude the possibility that this manipulation led to a specific evaluation of outcomes consistent with ARC-Outcome-bonus. However, we argue that the outcome-bonus model describes a reasonable mechanism to incorporate advice, even when advisors are not rewarded. According to ARC-Outcome-bonus, social information will especially influence behavior in difficult learning situations where the difference in payoffs between available options is smaller than the additional reinforcement through advice following. Additionally this model will, for good advice, also speed up learning when the payoff difference between options is already large. By contrast, the second-best model, ARC-Certainty will increase the probability of choosing the recommended option only when learning is difficult. Assuming additionally that people usually receive good advice or quickly identify bad advisors (Yaniv, 2004a), this suggests that integrating individual and social information according to ARC-Outcome-bonus would be more adaptive than according to ARC-Certainty, because it speeds up learning even when learning is not particularly difficult.

The specific situation in which we examined social learning is characterized by several aspects other than the formulation of advice and advisor incentives: participants were always only informed about the payoffs from chosen options; they received social information only once; and the social information was given as explicit advice instead of being provided as the opportunity to observe others' choices. We argue that the ARC models,

especially ARC-Outcome-bonus, can be applied to different situations with minor modifications. First, when participants are informed about forgone payoffs, ARC-Outcome-bonus could simply be modified to allow for the updating of nonchosen options, for instance, as described by Camerer and Ho (1999). An interesting prediction is that when forgone payoffs are also used to update propensities, the impact of advice on the learning process should be smaller. ARC-Outcome-bonus could also be used to model ongoing advice in every trial by adding a constant to payoffs when reinforcing the currently recommended options, just as we did for a single piece of advice. Finally, observational learning might also be modeled in the ARC framework by assuming that the options chosen by the majority receive additional reinforcements, or by adding reinforcement to options proportional to the frequency of others who were observed choosing these options (for an implementation of such a model, see McElreath et al., 2005). However, we expect the influence of advice to be stronger than the influence of observed choices (Gonzalez, 1994; Gonzalez & Tversky, 1990). Finally, when decision makers are informed about others' choices and payoffs, choice options could be reinforced by their own and by others' (discounted) payoffs. In sum, the ARC-Outcome-bonus social learning model that we propose can be applied to various social learning situations and we consider it an exciting endeavor to explore the model in these contexts.

8. Conclusion

At the outset of this article, we suggested that many decisions people make are based on their own experiences and the advice from other individuals. Of course, consumers' decisions or investors' portfolio allocations can be more complex than the repeated choice task we examined. Nevertheless, links to real-life decisions can be proposed from our paradigm of one-time advice prior to repeated choices from experience. For instance, the fact that consumers consistently choose expensive brand-name products when equally good but cheaper store brands are available could be explained by different evaluations of recommended and nonrecommended options (McClure et al., 2004). Similarly, the prominence of mutual funds in the presence of generally more successful index funds might be caused by banks' advice to buy mutual funds.

More generally, the research presented here suggests the following insights of practical relevance. First, our results indicate that people *combine* reinforcement and advice to make choices. Only a minority of participants in Experiment 1 relied exclusively on advice. Hence, advisors who want their advice to be followed should ensure that the recommended behavior is also associated with some immediate reward. Second, our results indicate that a one-time recommendation has a long-lasting influence on behavior. Thus, generalizing from our findings, repeated advice seems not to be necessary to guide behavior in a particular direction. Combining these two insights, successful advice for repeated decisions could focus on a single convincing recommendation and the association of immediate reward for the desired behavior, and less on repeated appeals to the decision maker. This observation may have implications in several domains where advice is typically given, for example, in

medical settings when adherence to recommended medical treatment affects possible outcomes.

The aim of this research was to investigate how advice influences learning in repeated decision making. In these studies people neither ignored nor blindly followed advice. Instead, they combined advice with their individual experience when making their choices. By integrating advice with an individual learning process people do not follow advice mindlessly but use the advice to accelerate the individual learning process, providing quicker solutions to decision problems.

Notes

1. We also assessed participants' risk preferences (cf. Holt & Laury, 2002), risk attitudes (e.g.,. Johnson, Wilke, & Weber, 2004; Weber, Blais, & Betz, 2002), and indecisiveness after the IGT. We found no meaningful correlations between these measures and adherence to advice or differences in model fits and parameters, so we do not report these data.

2. Yechiam and Busemeyer (2005) proposed a similar model that makes use of a utility function to transform received payoffs, which, owing to two additional parameters for gains and losses, makes their model more complex, and it employs a choice rule that increases sensitivity as a function of time. We also tested a one-parameter utility function and time-dependent sensitivity. As these more complex models did not achieve a substantially better fit, we only report on the results of the simpler models. The described RL model differs from Erev and Roth's (1998) model by assuming zero initial propensities, allowing negative propensities, and using an exponential choice rule with a sensitivity parameter instead of a simple proportional choice rule.

3. For instance, the vector of choice probabilities .4, .13, .13, .13 or .36, .36, .14, .14 have standard deviations of .133 and .127, respectively.

4. For instance, when using participants' observed choices to update propensities, the model can achieve a good fit for participants with alternating streaks of the same choice by setting the decay parameter to 1. In this case, propensities will always be zero, except for the option chosen in the last trial. Accordingly, the model would always predict that a participant repeats his or her choice of the last trial (given payoffs are positive) and, hence, would achieve a good fit without actually describing a learning process.

5. Note that we used the parameters obtained from the fitting procedure as described above. While the parameters were fitted to describe the choice of all decks, Figs. 5 and 6 show, for reasons of clarity, choice proportions for the decks that are relevant for the respective analyses.

6. As the influence of social learning in ARC-Initial and ARC-Outcome-bonus is defined as a function of the expected payoff of the best option in the set, the parameters in Experiment 1 can be applied without scaling them.

7. Participants were not made aware of this manipulation. This should not be considered as deception because receivers were not instructed that payoff distributions were stationary.

Acknowledgment

References

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

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*, 215–233.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, *104*, 371–398.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, *121*, 177–194.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, *14*, 253–262.

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*, 171–189.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. Oxford, England: Wiley.

Camerer, C. F., & Ho, T. H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*, 827–874.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, *104*, 137–188.

Celen, B., Kariv, S., & Schotter, A. (2005). *An experimental test of advice and social learning*. New York: Columbia Business School.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *Neuroimage*, *23*, 744–751.

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*, 951–978.

Erdem, T., & Keane, M. P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, *15*, 1–20.

Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*, *105*, 280–298.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*, 912–931.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, *88*, 848–881.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 94–107.

Estes, W. K. (1962). Learning theory. *Annual Review of Psychology*, *13*, 107–144.

Estes, W. K., & Maddox, W. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*, 403–408.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.

Gans, N., Knox, G., & Croson, R. (2007). Simple models of discrete choice and their performance in bandit experiments. *Manufacturing and Service Operations Management*, *9*, 383–408.

Gibson, S. K. (2004). Social learning (cognitive) theory and implications for human resource development. *Advances in Developing Human Resources*, *6*, 193–210.

Goetzmann, W. N., & Massa, M. (2002). Daily momentum and contrarian behavior of index fund investors. *Journal of Financial & Quantitative Analysis*, *37*, 375–389.

Gonzalez, R. (1994). When words speak louder than actions: Another's evaluations can appear more diagnostic than their decisions. *Organizational Behavior & Human Decision Processes*, *58*, 214–245.

Gonzalez, R., & Tversky, A. (1990). The impact of others' choice on investment decisions. In K. Borcherding, O. I. Larichev & D. M. Messick (Eds.), *Contemporary issues in decision making* (pp. 367–373). Oxford, England: North-Holland.

Hanaki, N., Sethi, R., Erev, I., & Peterhansl, A. (2005). Learning strategies. *Journal of Economic Behavior & Organization*, *56*, 523–542.

Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior*, *19*, 215–241.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, *12*, 123–135.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534–539.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*, 1644–1655.

Hutchinson, W. J., & Meyer, R. J. (1994). Dynamic decision making: Optimal policies and actual behavior in sequential choice problems. *Marketing Letters*, *5*, 369–382.

Johnson, J. G., Wilke, A., & Weber, E. U. (2004). Beyond a trait view of risk-taking: A domain-specific scale measuring risk perceptions, expected benefits, and perceived risk attitude in German-speaking populations. *Polish Psychological Bulletin*, *35*, 153–163.

Jungermann, H., & Fischer, K. (2005). Using expertise and experience for giving and taking advice. In T. Betsch & S. Haberstroh (Eds.), *The routines of decision making* (pp. 157–173). Mahawah, NJ: Erlbaum.

Kameda, T., & Nakanishi, D. (2002). Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evolution and Human Behavior*, *23*, 373–393.

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

Laland, K. N. (2001). Imitation, social learning, and preparedness as mechanisms of bounded rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 233–247). Cambridge, MA: MIT Press.

Lockhart, R. S. (2001). Memory retrieval. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 9613–9618). Oxford, England: Elsevier.

Luan, S., Sorkin, R. D., & Itzkowitz, J. (2004). Weighting information from outside sources: A biased process. *Journal of Behavioral Decision Making*, *17*, 95–116.

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 16075–16080.

March, J. G. (1996). Learning to be risk averse. *Psychological Review*, *103*, 309–319.

McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*, *44*, 379–387.

McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., Edsten, E., Efferson, C. & Paciotti, B. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior*, *26*, 483–508.

Meyer, R. J., & Shi, Y. (1995). Sequential choice under ambiguity—Intuitive solutions to the armed-bandit problem. *Management Science*, *41*, 817–834.

Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. Westport, CT: Greenwood Press.

Murray, F. S. (1971). Multiple probable situation: A study of a five one-armed bandit problem. *Psychonomic Science*, *22*, 247–249.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Rieskamp, J. (2006a). Positive and negative recency effects in retirement savings decisions. *Journal of Experimental Psychology: Applied*, *12*, 233–250.

Rieskamp, J. (2006b). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 1355–1370.

Rieskamp, J., Busemeyer, J. R., & Laine, T. H. (2003). How do people learn to allocate resources? Comparing two learning theories. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *29*, 1066–1081.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207–236.

Roider, A., Drehmann, M., & Oechssler, J. (2007). Herding with and without payoff externalities—An Internet experiment. *International Journal of Industrial Organization*, *25*, 391–415.

Rosenthal, T. L., & Zimmerman, B. J. (1978). *Social learning and cognition*. New York: Academic Press.

Schlag, K. H. (1998). Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, *78*, 130–156.

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

Schotter, A., & Sopher, B. (2003). Social learning and coordination conventions in intergenerational games: An experimental study. *Journal of Political Economy*, *111*, 498–529.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*, 233–250.

Simon, H. A. (1990). A mechanism for social selection and successful altruism. *Science*, *250*, 1665–1668.

Stahl, D. O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, *16*, 303–330.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*, 101–118.

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*, 263–290.

Yaniv, I. (2004a). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*, 75–78.

Yaniv, I. (2004b). Receiving other people's advice: Influence and benefit. *Organizational Behavior & Human Decision Processes*, *93*, 1–13.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281.

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*, 387–402.

Zimmerman, B. J., & Rosenthal, T. L. (1974). Observational learning of rule-governed behavior by children. *Psychological Bulletin*, *81*, 29–42.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*, 41–61.