CHAPTER 4

The Docile Learner:

How People Combine Advice and Reinforcement to Make Good Choices

Chapters 2 and 3 explained cooperative behavior in groups, by assuming that people use reciprocity as a decision rule. A problem for approaches which assume that decision rules determine people's behavior is that they have to explain how people learn and choose among decision rules. Different solutions have been proposed to this problem. In the tradition of research of the adaptive decision maker (Payne et al., 1993), it is assumed that decision makers trade off accuracy and effort to choose the rule that is most efficient given their cognitive capacities. March (1996) suggested an appropriateness framework, which assumes that decision makers use cues in the decision environment to decide which decision rule is (socially) appropriate. A third approach, individual learning, has recently gained prominence. For instance Rieskamp and Otto (submitted for publication) have proposed a strategy selection theory, which assumes that a reinforcement learning process describes how people learn among cognitive strategies. The aim of this chapter is to propose social learning as an additional account for strategy learning.

While the ultimate goal is to explain how people learn among decision rules, this chapter builds the foundation for this goal, by examining social learning among simpler choice options in the multi-armed bandit paradigm. Strategy learning is a complex process where decision makers first have to acquire the skill to apply a decision strategy and then learn which of the available strategies performs best in a specific task (Shrager & Siegler, 1998; Siegler & Araya, 2005; Siegler & Chen, 2002). As the social learning models I will examine are concerned with the second part, identifying the best option from a set of options, I will examine social learning in the multi-armed bandit paradigm, where people receive decision strategies as advice which are easy to use.

4.1 Introduction

Many decisions are made in a social context, where decision makers can observe others' decisions, or can seek and receive advice from other people. Accordingly, it has frequently been argued that we learn from others what to choose or how to make decisions (e.g. Bandura, 1977; Joseph Henrich & McElreath, 2003; Laland, 2001; A. Schotter & Sopher, 2003; Simon, 1955). Social information seems especially valuable in situations of uncertainty (Festinger, 1954), for instance, when the decision maker has little knowledge about the judgmental domain, when outcomes of choice options seem similar, or when information about objects or choice options

has to be accessed through one's own experience. Indeed, it has recently been argued that in many real-life situations, information about objects or choice options is not available in the form of attribute lists. Instead, people make decisions based on experienced consequences of their decisions and the involved learning process (Hertwig, Barron, Weber, & Erev, 2004). In the present article, I argue that in decision situations characterized by uncertainty and incomplete knowledge, people do not only make their decisions based on individual learning, but additionally based on others' advice. The main goal of the present article consists of exploring to what extent, and in which way, individual learning processes in repeated choice situations are influenced by the advice of others.

The classical paradigm to examine choices from options with uncertain payoffs is the multi-armed bandit paradigm, in which, in analogy to choice between different slot machines, decision makers choose between two or more options, of which at least one has an unknown payoff distribution (Sutton & Barto, 1998).[21] Because in this paradigm the decision maker is usually only informed about the payoff of the chosen option, he or she faces the problem of exploration versus exploitation. On the one hand, the decision maker wants to explore the options to find the one that provides the highest payoff. On the other hand, the decision maker wants to exploit his or her knowledge by always choosing the option with the highest estimated payoff. How decision makers trade-off exploitation and exploration in multi-armed bandits has been studied extensively (e.g. Gans, Knox, & Croson, 2004; e.g., Hutchinson & Meyer, 1994; Meyer & Shi, 1995; Murray, 1971; Vulkan, 2000). A prominent version of the multi-armed bandit problem is the Iowa Gambling Task (IGT, Bechara, Damasio, Damasio, & Anderson, 1994), in which decision makers repeatedly choose cards from four options (card decks) with different expected payoffs (for a review, see Maia & McClelland, 2004).

The optimal choice strategy in the multi-armed bandit problem is the Gittins Index strategy, which prescribes to choose the option with the highest Gittins Index (Gittins, 1989). The Gittins Index of an option is the sum of its expected payoff (which is updated according to Bayes' rule) and the increase of (discounted) future payoffs that can result from additional information gained by experimenting with that option. However, the Gittins Index is very complex to calculate, and is generally not assumed to be a descriptive model. Instead, learning in repeated choice tasks or decisions from experience has been modeled as having only limited demands on the decision maker's cognitive abilities. A common result of these studies is that

[21] I refer to *choices* from experience. Others have examined how covariations between attributes are assessed when objects are presented trial by trial versus in table form (see Allan, 1993; Kao & Wasserman, 1993; Shanks, 1991; Ward & Jenkins, 1965).

models implementing the law-of-effect (Thorndike, 1927) usually describe the learning process best.

Yechiam and Busemeyer (2005) have reviewed learning models for the IGT, which all assume initial expectancies for options that are updated on the basis of learning. First, the models can be distinguished as interference versus decay models, with the first updating only the expectancies of chosen options and the latter additionally reducing the expectancies of all options as a function of time. Second, the models can be distinguished as using either a maximizing or probabilistic choice rule. With a maximizing choice rule, a model always chooses the option with the highest expectancy, and explores alternative options with a constant probability. With a probabilistic choice rule, options are chosen probabilistically as a function of the options' expectancies. In addition to the learning models, Busemeyer and Stout (2002) have examined alternative models, the Strategy-Switching Heuristic and the Bayesian-Expected Utility Model, but these models described choices in the IGT less adequately than learning models. Modeling behavior in a two-armed bandit task, Gans et al. (2004) proposed a class of more complex models, implementing the Gittins Index and a class of cognitively less demanding models. Gans et al. used the Gittins Index strategy as one descriptive model. Their Myopic Model also uses the Gittins Index strategy, but calculates the Gittins Index of choice options based on a shortened (instead of an infinite) time horizon, and their Simple Model further simplifies the Myopic Model by assuming that players only distinguish between good and bad decks (i.e., they do not calculate expected payoffs from choosing a deck). The other, cognitively less demanding, models that Gans et al. proposed were Last-*n*, which calculates the expected payoff of a deck based on the last *n* payoffs from an option, Hot Hand, which assumes that options that were good recently will be good in the future, and an Exponential Smoothing, which has a functional form similar to simple reinforcement learning models (e.g., Erev & Roth, 1998). Erev and Barron (2005) proposed the RELACS Model to explain learning when decisions are made from experience (see, e.g., Barron & Erev, 2003; Hertwig et al., 2004; E. U. Weber, Shafir, & Blais, 2004) and for the probability learning task. RELACS assumes that a reinforcement learning process describes how decision makers learn among three different choice strategies. The first choice strategy, *fast best reply*, chooses the best response to recent outcomes, the second choice strategy, *loss aversion* and *case base reasoning*, avoids checks if options have high losses and searches for patterns in the payoff stream, and the third choice strategy, *slow best reply*, chooses options that were best over a long period of time.

A feature common to all models described above is that they assume learning, based exclusively on individually experienced outcomes, that is, the initial evaluation and the updating

process for all choice options are assumed to be identical. However, decision makers are often influenced by other people when making choices. Theories of social learning describe how people use social information in decision making and learning. Bandura's (1977) prominent social learning theory assumes that higher order cognitive processes and reinforcement processes are necessary to account for social learning. More specifically, the theory assumes that people learn simple behavior and complex concepts by observation and cognitive modeling, without imitating the role model and without being reinforced (Rosenthal & Zimmerman, 1978; Zimmerman & Rosenthal, 1974). However, reinforcement (personal or vicarious) is still important since it determines whether the learned behavior will actually be implemented. Recently, more specific computational social learning theories have been proposed. Inspired by the theory of the gene-culture evolution of Boyd and Richerson (1985), McElreath et al. (2004) proposed a model of imitation learning that combines individual learning with social learning by assuming that a choice option is reinforced through received payoff and through the observation that others choose that option. Apesteguia, Huck and Oechssler (2003) examined imitation behavior of interdependent individuals. They compared a model from Vega-Redondo (1997) in which decision makers observe their immediate competitors and imitate the behavior of the single most successful competitor, and a model from Schlag (1998; 1999) in which decision makers observe players in the same role (who are not competitors) and imitate other players dependent on how much higher their payoff is compared to their own payoff. The last group of social learning theories describes how individuals seek and integrate advice. Budescu and Rantilla (2000) describe how decision makers integrate expert opinions with a model that weights experts advice according to the amount of information that advisors had available. Yaniv and colleagues examined how decision makers integrate advice they received to update their own numerical judgments. Yaniv and Kleinberger (2000) found that decision makers put too little weight on others' advice, given the accuracy of advice receivers and advisors, whereby more knowledgeable decision makers discount advice more than the less knowledgeable (Yaniv, 2004b). Regarding the integration of advice, they found that lower weights were given to advices distant from their own initial estimate and to outliers in a distribution of advice values (Harries, Yaniv, & Harvey, 2004). Importantly, Yaniv (2004a; 2004b) points out that advice (from independent decision makers) generally improves performance. Luan, Sorkin and Itzkowitz (2004; submitted for publication) tested the use of advice in signal detection tasks and found that players are sensitive to the quality of advice—they put higher weights on better advisors—and that they search advice in an adaptive fashion when they can decide if, and from whom, to seek advice.

While these models of social learning explain how social information is used in the decision situations they were developed for, I aim to provide a complement to these theories that explains the use of single advice in repeated choice situations. More specifically, existing models of advice-seeking and integration describe how advice is used for single judgments and not for repeated choices, and models of imitation describe imitation when others can be observed before each choice, but not when one observation is followed by repeated choices. To fill this gap, this article proposes and tests models, explicitly modeling how decision makers use individual experience and one-time advice when making repeated choices.

The article is structured as follows. In the next section, I describe the paradigm I use to examine advice following in repeated choice tasks, and report on the results in Experiment 4.1 which examined if decision makers would follow advice. I then illustrate models that aim to describe advice receivers' behavior and report how well they describe behavior. A second experiment then tests, among the three models accounting best for advice receivers choices. I conclude with a general discussion.

4.2 Experiment 4.1

Experiment 4.1 examined how social learning influences choices in the Iowa Gambling Task (IGT, Bechara et al., 1994). In the IGT, participants receive an initial endowment (10 euro in Experiment 4.1), and then choose cards from four different decks (A, B, C, D). Whenever a participant chooses deck A or B, he or she receives a reward of .5 euro, when he or she chooses deck C or D, he or she receives .25 euro. Participants sometimes also incur a loss when choosing a deck. Losses from choosing decks C and D (henceforth "good decks") are moderate, so that the expected payoff from those decks after 100 trials is 12.5 euro. Losses from decks A and B (henceforth "bad decks") are so large that the expected payoff for these decks is –12.5 euro. The differences between decks with the same expected payoff is that one deck has relatively frequent, but low losses (low variance), whereas the other deck has rare, but high losses (high variance).
The payoff schedule I used is identical to the schedule introduced by Bechara et al. (1994). A crucial property of this schedule is that losses from the bad decks (A and B) occur relatively late, so that the bad decks initially seem to be better. When choosing in the IGT, participants usually need at least 20 trials to learn which decks allow to earn money, and after that still sometimes choose one of the bad decks (Maia & McClelland, 2004). The question the first experiment addresses is whether social learning can improve participants' performance by helping them to detect the good decks earlier and also by increasing the likelihood to choose good decks later in the task. A useful 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 adhere to the advised deck in the presence of an equally attractive alternative (henceforth "corresponding deck").

4.2.1 Method

4.2.1.1 Design

To examine the effect of social learning, participants performed the IGT with, and without, advice. Independent participants in condition one performed the IGT without receiving or giving advice. Participants in condition two, advisors, performed the IGT without receiving advice, then chose one of several predetermined advices for another participant, and finally performed the IGT again. Participants in condition three, receivers, received advice from an advisor and then performed the IGT.[22]

4.2.1.2 Participants and Procedure

In the experiment, 90 participants, mostly students from the Free University of Berlin (54% women with a mean age of 25 years), were randomly assigned to the three conditions.

In the independent learning condition, participants were instructed that they are 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 a 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 are associated with which decks.

To inform all 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 payoffs from the decks or the order of the payoffs within a deck. To further clarify the stochastic nature, the last 20 participants in each condition were asked to imagine that they choose from actual card decks. As behavior (i.e., frequency to choose good decks, and adherence to advice of advice receiver) was the same for all participants, I will not distinguish between the first 10 and last 20 participants in the conditions.

After the introduction of the task, participants were told that they would start the task with an initial endowment of 10 euro, they were reminded that their show-up payment was 5 euro, and that they would receive their final account balance minus the 10 euro initial endowment as a

[22] I also assessed the participants' risk preference (cf. Holt & Laury, 2002), risk attitudes (cf. Johnson, Wilke, & Weber, 2004; E. U. Weber, Blais, & Betz, 2002), and indecisiveness after the IGT. As I found no meaningful correlations between these measures and adherence to advice or models' fits and parameters, I will not report on these data.

variable payment. In case the final account balance was negative, participants still received the show-up payment (they learned this only after the experiment).

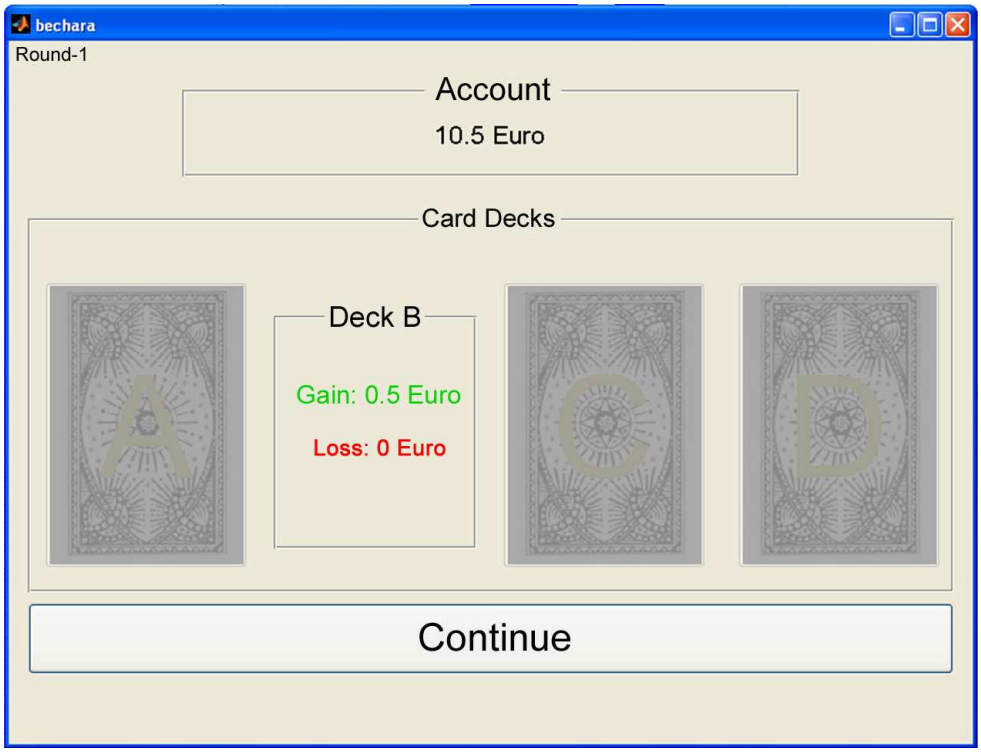

Figure 4.1. Graphical user interface for participants in Experiments 1 and 2. Participants chose decks by clicking on the decks.

Finally, participants were briefly instructed about the graphical user interface (see Figure 4.1) used to conduct the experiment. After choosing a card 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. To continue with the next trial, participants had to click the "continue" button. The minimum time interval between two choices was fixed to 3 seconds, no upper limit was fixed.

Advisors received the same information as independent decision makers, plus additional information about their role as advisors. Specifically, they were first informed that they would choose an advice for another participant, who would perform the identical task, after completing their decision-making task. In order to be able to evaluate if receivers actually follow advice, a set of feasible advices was predefined. Advisors could choose among four advice strategies which were "choose always from deck A" (or "B", or "C", or "D"). The feasible advices were 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 indicated his or her advice on a form, which was then given to the receiver. To motivate advisors and to make them credible to advice receivers, they received, additionally to the payoff from their own choices, 50% of the receivers' payoffs.

Receivers received first the instructions of the IGT, and then the form with the feasible advices, one of which was marked as advice by an advisor in the same session. Receivers were informed that the advisor had sufficient experience with the task to give good advice, and that the advisor would receive a payment equivalent to 50% of the receiver's variable payment from the IGT. As in the other conditions, receivers' variable payment depended on their performance in the IGT.

Experiments were conducted in sessions with two to six participants. All participants completed the experiment in separate cubicles, thus preventing any interaction.

4.2.2 Results From Experiment 4.1

4.2.2.1 Choices and Performance

Participants earned, on average, 5.02 euro (SD = 5.31) in the IGT. Independent decision makers chose one of the two good decks, on average, in 62% (SD = 14%) of all 100 trials, which is less than the proportion of 73% (SD = 2%) with 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 78% (SD = 17), $t(29) = 2.54$, $p < .001$, $d = 0.66$. Receivers chose one of the good decks across their 100 trials more frequently than advisors in their first 100 trials, $t(29) = 3.11$, $p = .003$, $d = 0.8$. The advisors chose one of the good decks in, on average, 59% (SD = 14) of the first 100 trials and in 73% in the last 100 trials, thus, they increased their performance significantly, $t(29) = 4.75$, $p < .001$, $d = 1.23$.

Figure 4.2 shows, for 100 trials of each condition, the proportion of participants who chose one of the two good decks. This proportion declined at the beginning (i.e., 10 to 20 trials) for all groups, with the exception of the advisors at the beginning of their second 100 trials.

Figure 4.2 also shows that receivers at the beginning of the task perform better than advisors in the beginning of their second 100 trials. However, starting at about trial 15, due to a strong increase of the advisors' performance, the receivers perform worse than the advisors, and only at the end of the 100 trials do both groups perform equally well again. In sum, advice generally improves performance, compared to inexperienced participants, with the advantage of being especially large in the first trials.

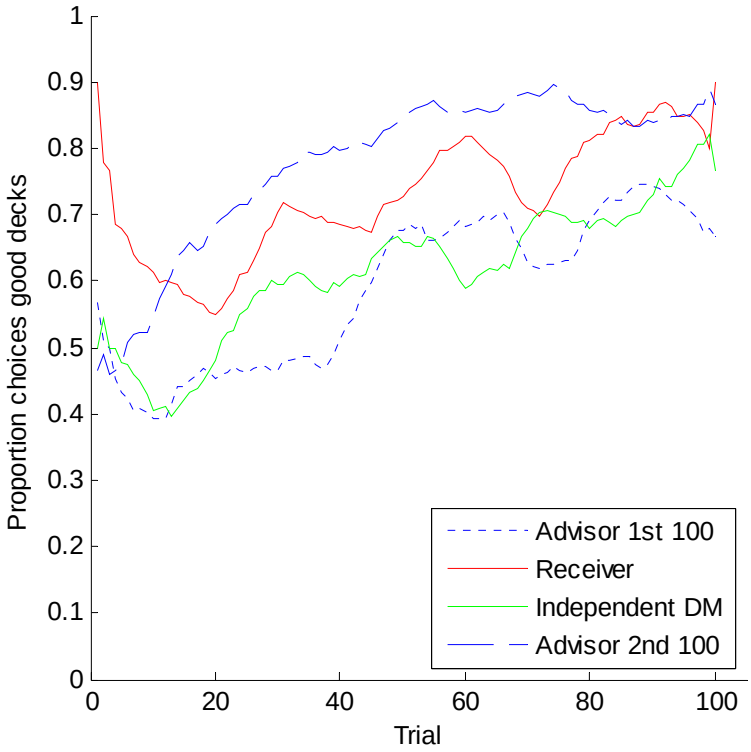

Figure 4.2. Proportion of participants who chose one of the two good decks (moving average, with a window size of 11).

4.2.2.2 Advice-giving and –following

The large majority of 28 participants (93%) in the role of the advisors gave good advice. From these 28 advisors, 19 advisors proposed to choose the good deck with a high payoff variance and 9 proposed to choose the good deck with a low payoff variance. The fact that the IGT contains two options with the same expected payoff allows us to test the influence of advice. If participants follow the advice, they will not only most likely choose a good deck, as one could also predict by individual learning, but in addition they should prefer the advised deck out of the two good decks.

Receivers chose a good deck with low variance when it was advised, on average, in 62% of the trials (*SD* = 9%), whereas the mean percentage was 10% (*SD* = 3%) when it was not advised. The mean percentage for a deck with high variance was 69% (*SD* = 9%) when it was advised, but 7% (*SD* = 3%) otherwise. Figure 4.3 shows the development of choice proportions over the 100 trials, and reveals that participants nearly always start by choosing the advised deck, but then quickly switch to choosing other decks. However, after approximately 20 trials, the percentage with which participants choose the advised deck increases again. I refer to this sequence as the adherence-exploration-adherence pattern. Altogether, the results show a strong influence of

advice on choices because receivers clearly prefer the advised deck to the corresponding deck with the same expected payoff.

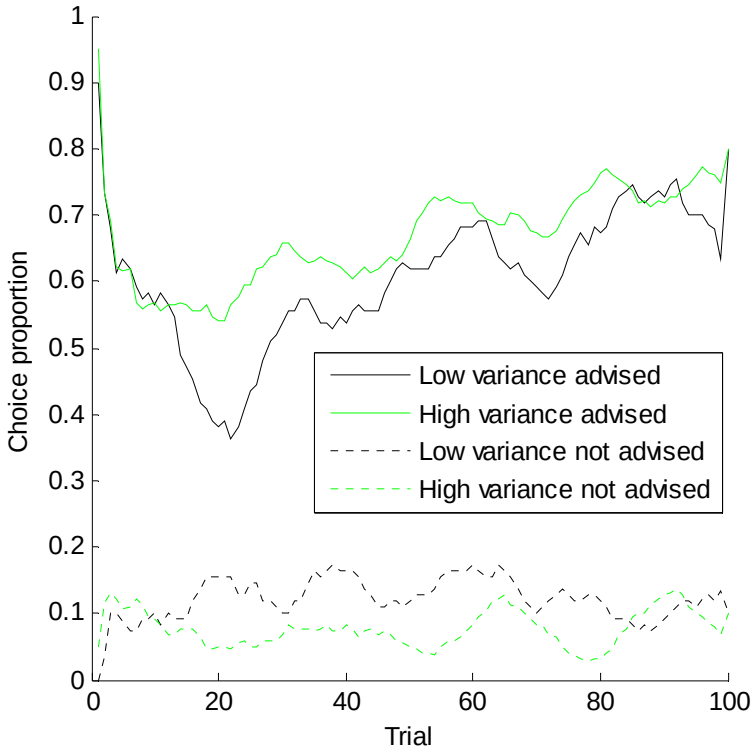

Figure 4.3. Proportion of participants who chose a deck conditional on advice (moving average, with a window size of 11). Note that decks with the same expected payoff do not need to sum up to one because participants can also choose from the two other decks.

The analysis of participants' choices and advices suggests that participants without advice learn to choose the good decks, receivers follow the advice they received, and advice gives receivers an advantage especially in the first choices. The decline and rebound of the probability with which receivers chose according to the advice suggests that they combined advice information with individual experience to determine which choices to make. The next section proposes computational models describing how specifically people combine advice and payoff information.

4.3 Models of Learning in Repeated Choice Tasks

Several types of models have been proposed to describe how individuals learn to choose among choice options. As reinforcement learning models have been most successful in describing people's choices (Busemeyer & Stout, 2002; Erev & Barron, 2005; Gans et al., 2004; Yechiam & Busemeyer, 2005), I restrict our examination to variants of simple reinforcement

learning models.[23] The learning models I propose are similar to the models suggested by Erev and Roth (Erev, 1998; Erev & Roth, 1998) for learning in normal form games, and by Busemeyer and colleagues (Busemeyer & Stout, 2002; Yechiam & Busemeyer, 2005) for learning in the IGT. Similar models have been used to explain learning in repeated decisions (Camerer & Ho, 1999b; Erev & Barron, 2005; Estes, 1962). All these models are extensions following early learning models by Bush and Mosteller (1955) and Estes (1950).

The decision problem consists of choosing among options *i* out of a set $S = \{1,2,...,n\}$ with *n* options. Before making a decision, the decision maker might receive an advice $A = \{1,...,m\}$ which is a subset of *S*. Generally, an advice can consist of one or several options. After choosing option *i* in trial *t*, the decision maker receives a payoff $\pi_t(i)$.

4.3.1 Individual Learning

Models of reinforcement learning assume that participants' choices are determined by the law-of-effect (Thorndike, 1927), that is, options which produced higher payoffs are chosen with a higher probability. Recently, reinforcement learning models have regained prominence to explain decision making in various domains (Camerer & Ho, 1998; Erev, 1998; Rieskamp & Otto, 2004; Stahl & Haruvy, 2002). Estes (1962) and Vulkan (2000) reviewed the application of learning models to repeated choice tasks, especially probability learning.

According to the individual Reinforcement Learning model (RL), the choice options have initial propensities before any choices are made. After choosing an option, the resulting payoff is used to update the option's propensity, which also decays with time. The probability to choose an option is a function of its propensity.

Formally, the initial propensity of a strategy before the first decision is $q1(i) = 0$. Independent of choices, the propensities of options decay with time. After choosing an option *i*, the propensity $q(i)$ of an option is updated with the received payoff, so that:

$$q_{t+1}(i) = (1-\phi) \cdot q_t(i) + r_t(i) \tag{4.1}$$

where ϕ is a free decay parameter determining the weight of past experiences in the updating process, and 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

[23] I also implemented a Bayesian updating model that incorporates advice and simpler Win-Stay Lose-Shift learning models. However, both model types performed consistently worse than the reinforcement models. Bayesian models have also been outperformed by reinforcement learning models in other tests (Busemeyer & Stout, 2002; Gans et al., 2004). Hence, in the interest of brevity, I omit their description.

$$p_t(i) = e^{\lambda \cdot q_t(i)} / \sum_{j=1}^{n} e^{\lambda \cdot q_t(j)}, \quad (4.2)$$

To capture the variability in participants' sensitivity to differences in propensities, the choice rule is augmented by a sensitivity parameter λ (cf. Yechiam & Busemeyer, 2005). I further assume that participants choose the advised option in their first trial. While this brings some social learning in RL, and makes it a nested model of the more complex social learning models, it makes it a stronger competitor of the social learning models.

The proposed learning model is similar to previously proposed learning models. The model by Yechiam and Busemeyer (2005) makes use of a utility function to transform received payoffs, which, due to two additional parameters for gains and losses, makes their model more complex, and it employs a choice rule which increases sensitivity as a function of time. I also tested with a (one parameter) utility function and time-dependent sensitivity. As these more complex models did not achieve a better fit, I only report on the results of the simpler models. The described reinforcement 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 ration rule.

4.3.2 Social Learning

A model of social learning needs to specify to what extent the learning process is influenced by individual experience and by information from other people. Hence, the following social learning models combine information from advice with an individual reinforcement learning process, and are, thereafter, named Advice Reinforcement Combination (ARC) models. The models incorporate advice by assuming that advice changes either the initial evaluation of choice options, the evaluation of payoffs from the advised options, the decay of propensities, or the choice rule of advised options. To specify the ARC models formally, I change the individual learning model, RL described above, by adding mechanisms, so that RL is a special case of the ARC models (i.e., it is a nested model). All ARC models assume that advice receivers always choose the advised option in the first trial. The motivation behind this assumption is that advice receivers will attempt to evaluate the advised option before exploring alternative options. Formally, the probability to choose an option in the first trial is $p(i|i \in A) = 1/m$ and $p(i|i \notin A) = 0$.

4.3.2.1 ARC-Initial

One way to introduce social information into the individual learning process is to assume that decision makers initially perceive advised options as more positive, compared to alternative nonadvised options. Such an assumption is reasonable because advisors usually have more knowledge than advice receivers (Jungermann & Fischer, 2005). Decision makers who expect

higher payoffs from the advised option and trust in the advisors' competence should start with the advised option and only slowly revise their initial judgments in the case of conflicting evidence. To model the assumption of an initial preference for the advised option and a slow revision process, I allow the initial propensity of the advised option to be higher than for options that were not advised. In a similar way, Camerer, Ho, and Chong (2002) and Hanaki, Sethi, Erev, and Peterhansl (2005) model the own experience, for which advice can be regarded as a substitute, with choice options, by defining initial propensities as a function of past payoffs from the options.

Formally, the initial propensities for ARC-Initial are defined as $q1(i|i \in A) = |\mu| \cdot \iota$ and $q1(i|i \notin A) = 0$, where ι is a free parameter determining the additional initial propensity of the advised strategy, and μ is the expected payoff from always choosing the best option in the set.

4.3.2.2 ARC-Reinforcement

Social information can also, instead of changing initial evaluations, influence the continuing evaluation of payoffs from the advised option, so that the consequences of advised options are perceived more positively, compared to the consequences of nonadvised options. This assumption is consistent with research on conformation biases, in the case of social learning implying that people believe in the judgment of the advice giver, and overvalue confirming information and undervalue disconfirming information (for a review see Nickerson, 1998). Empirical evidence suggests that the confirmation bias can influence choices in repeated choice tasks. For instance, Betsch, Haberstroh, Glockner, Haar, and Fiedler (2001) first induced people to make a particular choice and then changed the environment so that the induced choice was no longer the best, and found that people overweigh confirmatory information and underweigh disconfirmatory information after the environment had changed. Moreover, Aronfreed (1969) has argued that imitating the behavior of others, in its own, is sufficient to generate affective reward. Thus, choosing the advised option might lead to a positive reinforcement, regardless of the real consequences of the choice. A generally more positive evaluation of outcomes from advised options can be implemented in the learning model by adding a constant to every payoff from the advised option. Formally, reinforcements for advised strategies 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 an advised option.

4.3.2.3 ARC-Decay

Advice cannot only influence the evaluation of outcomes from choice options but also the decay of propensities for options. Specifically, it can be assumed that advised options have, due

to their prominence, stronger memory traces. Experimental tests of trace dependent theories of memory showed that memories with stronger traces are easier to retrieve (Lockhart, 2001). Hence, it should be easier to retrieve information about the past performance of the advised options, which will then have a greater impact in the updating process, compared to the past performance of nonadvised options. I implement this assumption by introducing a second decay parameter for the advised option, which is assumed to be lower than the decay parameter for the other options. Formally, δ is the free decay parameter for the advised option, while ϕ remains the decay parameter for the nonadvised strategy. The important implication of the ARC-Decay model is that the accumulation of (negative or positive) propensities will be faster, and their reduction slower, for the advised option than for alternative options. Formally, the different decay process is implemented by modifying Equation 4.1 to:

$$q_{t+1}(i) = \begin{cases} i \notin A & \rightarrow \ (1-\phi)\cdot q_t(i) + r_t(i) \\ i \in A & \rightarrow \ (1-\delta)\cdot q_t(i) + r_t(i) \end{cases}. \tag{4.3}$$

4.3.2.4 ARC-Choice

The social learning models presented so far assume that social learning directly influences the learning mechanism. An alternative possibility is that learning is not influenced by social information, but by choices that are made based on one's own experience. This reasoning is based on Festinger's (1954) insight that people rely on social information when making judgments, especially when they are uncertain about their own judgment. This argument has also been used to model the evolution of social learning. Henrich and Boyd (1998) and later Kameda and Nakanishi (2002; 2003) showed that in the evolutionary competition between social learners, who decide individually when they are certain and copy the majority when they are uncertain, and individual learners, a stable proportion of social learners can be maintained. I implement the intuition that people rely on social information when they are uncertain by assuming that decision makers choose according to propensities when the variance of propensities is high and choose the advised option when the variance of propensities is low. Formally, I model reliance on the advice contingent on the variance of choice probabilities because this makes the model parameter independent of the magnitude of payoffs. Specifically, probabilities are modified, after they were calculated with Equation 4.2, according to following function:

$$p(i) = \begin{cases} \sigma[p(i)] < \tau \ \vee i \in A & \rightarrow \ 1/m \\ \sigma[p(i)] < \tau \ \vee i \notin A & \rightarrow \ 0 \\ \sigma[p(i)] \geq \tau & \rightarrow \ p \end{cases}, \tag{4.4}$$

where $\sigma[p(i)]$ is the standard deviation of the choice probabilities for Equation 4.2 and τ is a free uncertainty parameter, which determines the threshold below which the advised option is chosen. *Comparison of social learning mechanisms.* I have proposed three different social learning mechanisms describing how advice could influence individuals' learning processes. Formally, these mechanisms were added to the individual learning model, so that three more complex learning models resulted. To compare the social learning models, I consider at which point in time social learning has its largest influence, how persistent the influence is over time, how learning proceeds if the advised option is not the best available option, and how the learning models describe learning in the loss domain. Table 4.1 summarizes the functions used to describe the ARC models and the RL model.

Table 4.1. *Learning and choice mechanisms in ARC.*

Mechanism	Individual learning (RL)	Difference between individual and social learning
First choice	$p_1(i) = \begin{cases} i \in A & \rightarrow 1/m \\ i \notin A & \rightarrow 0 \end{cases}$	*No difference*
Initial attraction	$q_1(i) = 0$	ARC-Initial: $q_1(i) = \begin{cases} i \in A & \rightarrow \lvert\mu\rvert \cdot \iota \\ i \notin A & \rightarrow 0 \end{cases}$
Rein-forcement	$r_t(i) = \pi_t(i)$	ARC-Reinforcem.: $r_t(i) = \begin{cases} i \in A & \rightarrow {}_t(i) + \lvert\mu\rvert \cdot \rho \\ i \notin A & \rightarrow \pi_t(i) \end{cases}$
Updating	$q_t(i) = q_t(i) \cdot (1-\phi) + r_t$	ARC-Decay: $q_t(i) = \begin{cases} i \in A & \rightarrow q_t(i) \cdot (1-\delta) + r_t \\ i \notin A & \rightarrow q_t(i) \cdot (1-\phi) + r_t \end{cases}$
Choice rule	$p(i) = e^{\lambda \cdot q(i)} / \sum_{j=1}^{n} e^{\lambda \cdot q(j)}$	ARC-Choice: $p(i) = \begin{cases} \sigma[p(i)] < \tau \vee i \in A & \rightarrow 1/m \\ \sigma[p(i)] < \tau \vee i \notin A & \rightarrow 0 \\ \sigma[p(i)] \geq \tau & \rightarrow p \end{cases}$

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

ARC-Initial assumes that advice mainly has an impact on the beginning of the learning process, and that social influence is not persistent because the decay of propensities consistently reduces the impact of the initial propensities. ARC-Initial predicts that, in the long run, advice receivers will learn to deviate from an advised option when a better alternative is available. The reason is that the impact of the higher initial propensity will diminish over time, so that individuals will experience that other options in fact produce better outcomes than the advised option (unless the initial propensity and the sensitivity parameters are extremely high). ARC-

Initial always influences the learning process in favor of the advised option, regardless of whether decisions are made in the gain or the loss domain.

Social learning, according to ARC-Reinforcement, is persistent and accumulates during learning, so that its impact is relatively small in the beginning, but increases thereafter. Due to the continuous additional reinforcement for advised options, decay does not reduce the influence of social information as it does for ARC-Initial. Instead, the influence of social learning increases with time. As in ARC-Initial, ARC-Reinforcement predicts that the choices of the advised options also increase in the loss domain. Differently than in ARC-Initial, ARC-Reinforcement predicts that advice receivers can adhere to the advised option in the presence of better alternatives since the additional reinforcement for this option make it appear better than the alternative options, provided that the advantage of the dominating option is not too large.

Social learning in ARC-Decay is persistent and can (in the gain domain) explain why advice receivers adhere to the advised option in the presence of better alternatives. In the domain of losses, ARC-Decay predicts that advice receivers will tend to avoid advised options because the slower decay for propensities of this option will maintain negative propensities longer, thus strengthening the advantage of alternative options for which negative propensities decay faster.

Social learning in ARC-Choice depends less on time and more on the similarity of two choice options. Generally, the more similar two choice options are, or the higher the variance of the choice options is, the more influence social information has. Nevertheless, social information is persistent because the advised option will also be preferred in later choices when the variance of choice options is low. ARC-Choice can also predict a preference for the advised option, in the gain and in the loss domain, when the variance of the options is high.

In sum, the comparison of the social learning models shows that they make different predictions independent of specific parameter values. Especially ARC-Reinforcement, ARC-Decay, and ARC-Choice can explain that decision makers keep choosing the advised option in the presence of a better alternative, and that advice directly influences later choices. Only ARC-Decay is consistent with faster deviation from advice when the expected payoff from the advised option is negative, independent if advice was good or bad. The next section examines how well the models describe the advice receivers' learning process and choices in Experiment 4.1.

4.4 Testing Models of Social Learning

As a first step, I had to estimate the models' parameters. For this purpose, different routes can be taken. First, model parameters can be estimated based on averaged data or on individual data. As recent comparisons of these two approaches showed that the estimation of model

parameters based on individual participants identifies true parameter values better (Estes & Maddox, 2005), I estimated model parameters for each participant. A second decision concerns the question of whether a model's prediction for a particular decision of a participant should be able to make use of the participants' actual past decisions. For instance, the propensities of the learning models for the second period can be updated based on the payoff from the choice predicted by the model in the first trial, or based on the payoff from the choice the participant actually made in the first trial. When making use of participants' past decisions a model will obtain a better fit than when the models' predictions are completely independent of participants' decisions since, in the former case, any wrong predictions will not enter the updating process, whereas, in the latter case, wrong predictions will effect subsequent predictions of the model. For this reason, I choose the more demanding approach, so that the models' parameters were estimated completely independent of the participants' behavior, and the models' predictions were determined only on the basis of the models' past predictions.[24] This approach should provide a more illuminative test of the models because it required the model to predict choices and the learning path to achieve a good fit.

Specifically, all models determine the probability with which an individual chooses an option, based on past choices and parameter values. I relied on the maximum likelihood estimation to find the best parameter values, that is, I searched for the parameter values that maximized the sum of the log likelihood of the observed behavior. The sum of the log likelihood is defined as $LL = \sum_{t=1}^{T} \ln(p_t(k))$, 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, every parameter combination which gives the chosen option only once a zero could never be selected as the best parameter combination. For this reason, I fixed the minimum choice probabilities in the fitting process to .001. To account for the probabilistic nature of the choice rules, choices were selected randomly according to propensities. In order to generate the average learning process for a set of parameter values, the models predictions for a particular set of parameter values were simulated 50 times, and the likelihood of the data, given the model, was determined based on the average probabilities for the choices over the 50 simulations.

[24] For instance, when using participants' observed choices to update propensities, the model can achieve a good fit for participants with long 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, achieve a good fit without actually describing a learning process.

The model parameters were constrained to $\phi \in [0,1]$ and $\delta \in [0,1]$ for the decay parameters, to $\lambda \in [-5,5]$ for the sensitivity parameter, to $\rho \in [0,10]$ for the additional reinforcement in ARC-Reinforcement, to $\iota \in [0,100]$ for the higher initial attraction in ARC-Initial, and to $\tau \in [0,.5]$ for the threshold in ARC-Choice, where .5 is the maximum standard deviation for a choice set with four options. To identify the best parameter values, I first performed a grid search, and then used the best five parameter sets of the grid search as start values for the simplex optimization algorithm (Nelder & Mead, 1965) to determine the best parameter values.

To evaluate the basic model performance, each model was compared to a statistical baseline model. The statistical baseline model is a three parameter model assuming that decision makers always choose the same option with the same probability, which is determined as the proportion of choices a participant made for that option over the 100 trials of the IGT. I accounted for differences in the models' complexity by determining the Akaike Information Criterion (AIC, see, e.g., Zucchini, 2000) for each model. I compared each model with the statistical baseline model by computing the difference of the AIC for the model and the statistical baseline model:

$$\Delta\text{AIC} = -2 \cdot [\text{LL}(\textit{model}) - \text{LL}(\textit{baseline})] - d \cdot 2, \tag{4.5}$$

with d as the difference in the number of parameters between the respective models and the statistical baseline model. The ΔAIC increases with a model's fit and decreases with the model's complexity, estimated by the model's number of parameters. The AIC can be used to compare non-nested models, although this comparison has to be interpreted cautiously since AIC does not take the model's functional form into account.

To evaluate the models, I examined first if they are better than the statistical baseline model, by examining the ΔAIC measurement. Table 4.2 shows that only the social learning models have, on average, positive ΔAICs, indicating that they perform better than the statistical baseline model (the same holds when one uses the Bayesian Information Criterion, which penalizes stronger for more parameters when $T > 7$). As a second step, I examined if the social learning models perform better than RL. Table 4.2, and the results of t-tests depicted in Table 4.3, show that all social learning models are, on average, better than the individual reinforcement learning model. Together with the previous finding that the advised deck was chosen more frequently than the corresponding deck, this result clearly supports the assumption that a social learning process describes decision-makers' choices better than pure individual learning.

Table 4.2. *Means (SD) for ΔAICs and parameter values for ARC models.*

Parameter or model fit	RL	ARC models			
		Initial	Reinforcement	Decay	Choice
Δ*AIC*	-7.5 (28.6)	3.73 (20.72)	8.7 (16.2)	6.52 (18.32)	6.67 (14.21)
Social learning	-	31.66 (36.69)	3.68 (3.48)	.16 (.25)	.17 (.16)
Decay	.12 (.21)	.16 (.30)	.51 (.38)	.34 (.3)	.36 (.34)
Sensitivity	3.03 (2.11)	2.16 (2.14)	2.86 (1.89)	2.55 (2.56)	2.85 (2.67)
RMSD					
Good deck	.062	.052	.056	.059	.062
Adherence	.102	.049	.034	.037	.041
All choices	.049	.043	.041	.043	.046

Note. Social learning parameters are additional initial attraction (ι) for ARC-Initial, additional reinforcement (ρ) for ARC-Reinforcement, separate decay (δ) for ARC-Decay, and standard deviation threshold (τ) for ARC-Choice. RMSD is the mean over all trials of the squared deviation between predicted and observed choice proportions on the group level. For "good choices," all choices of a good deck were considered, for "adherence," all choices in which the advised or the corresponding deck was chosen were considered, and for "all choices," all choices were considered.

To examine if one social learning model outperforms the other ARC models, I conducted t-tests comparing ΔAICs, for which the results are depicted in Table 4.3. A comparison of the social learning models shows that while ARC-Reinforcement seems to be the best model, it is significantly only when compared to ARC-Initial. The average fit of ARC-Reinforcement is better than for ARC-Decay and ARC-Choice, but the difference has only a small effect size. In sum, the comparison of the average model fits show that social learning explains participants' choices better than a statistical baseline model and better than the individual reinforcement learning model. Among the social learning models, ARC-Reinforcement has the best average ΔAIC values and also the best RMSD over all choices (see Table 4.2). However, it is not significantly better than ARC-Decay or ARC-Choices.

Table 4.3. *T-tests between models.*

	ARC-Initial	ARC-Reinforcement	ARC-Decay	ARC-Uncertain
RL	t = 3.34, p = 0, d = .44	t = 4.25, p = 0, d = .66	t = 4.98, p = 0, d = .56	t = 3.57, p = 0, d = .60
Initial		t = 2.94, p = .01, d = .27	t = 1.27, p = .21, d = .14	t = 1.38, p = .18, d = .17
Reinforcement			t = -1.33, p = .19, d = .13	t = -1.97, p = .06, d = .13
Decay				t = .09, p = .93, d = .01

Note. For all tests df = 29. When t statistics are negative, the row model is better than the column model.

Considering the social learning parameters, ι for ARC-Initial is 31.66—that is, the initial attraction for the advised models is approximately 32 times the average payoff from a good deck (.125 cents), ι for ARC-Reinforcement is 3.68—that is, every reinforcement from an advised deck received an additional 3.68 times the average payoff from a good deck, and δ for ARC-Decay is .16, therefore, clearly lower than the decay rate of .49 for options that were not advised. In ARC-Choice, participants chose the advised option, on average, when the standard deviation of the choice probabilities was below τ = .17. For instance, the vector of choice probabilities .55, .183, .183, .183 or .42, .42, .08, .08 have a standard deviation of .173.

An alternative way to compare the models is to examine, for how many participants a specific model is the best model. ARC-Decay describes most participants (10) best, followed by ARC-Reinforcement (7.67[25]), ARC-Initial and ARC-Choice (each 4.67), and RL (3), thus confirming the superiority of ARC-Reinforcement and ARC-Decay. Allocating participants to the three best models according to ΔAICs, ARC-Decay is the best for 14 participants, ARC-Reinforcement for 10.33, and ARC-Choice for 4.67. An examination of the raw data also suggested that participants used advice differently. Specifically, five participants made all 100 choices according to the advice, a behavior that seems most in line with ARC-Initial, but can also be modeled by ARC-Reinforcement and ARC-Choice.

Apart from comparing the model fits, one can query whether the models can predict characteristic patterns of choices over time. Figures 2 and 3 show that advice receivers first

[25] If n models had the best ΔAIC for a participant, 1/n participant was assigned to each.

follow the advice, thereafter deviate from it (e.g., to explore alternative options), and finally follow the advice again—they show an adherence-exploration-adherence choice pattern. Can the ARC models predict this pattern? Figures 5 and 6 compare the average observed choice patterns with the average choice patterns predicted by the ARC models. The predictions of the ARC models 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. Figure 4.5 compares with which probability real and simulated participants chose one of the two good decks in the IGT. To evaluate the correspondence of simulated and observed choices, I calculated the root mean square deviation (RMSD, see Table 4.2) between predicted and observed average probabilities on the group level. The RMSDs for "good decks" in Table 4.2 indicate that all social learning models predict the proportion of choices of good decks similarly well (see also Figure 4.4).

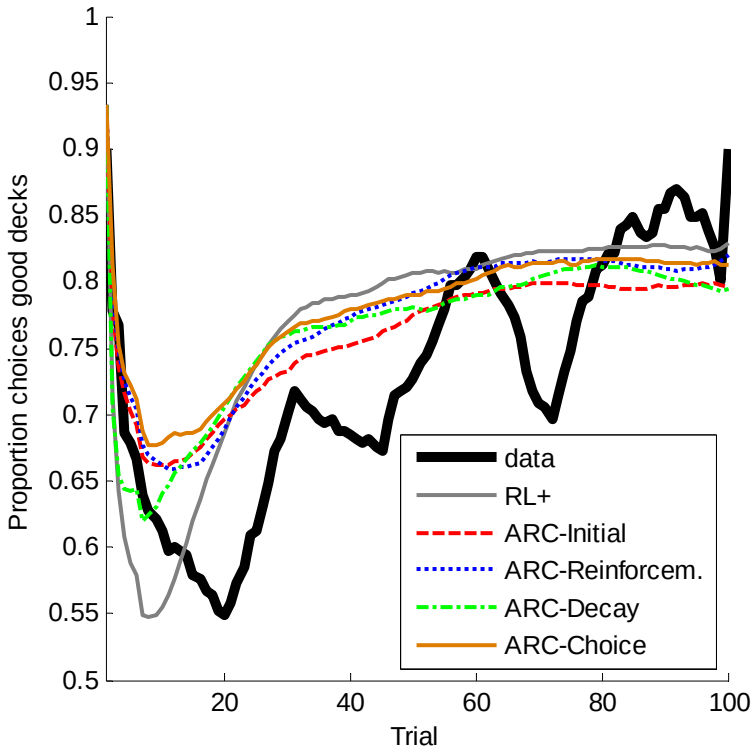

Figure 4.4. Proportion of participants choosing a good deck. Comparison of observed and simulated choices.

The picture is different when one adds the information, if a deck was advised or not. Figure 4.6 compares with which probability real and simulated participants chose a good or bad deck, given that it was advised or not. For instance, when a participant received the advice to choose the good deck C, I calculated with which probability this deck was chosen and with which

probability the corresponding (equally) good deck D was chosen. Figure 4.6 and the RMSDs in Table 4.2 show that only the ARC models that implement social learning—especially ARC-Reinforcement, ARC-Decay, and ARC-Choice—can account for the adherence-exploration-adherence pattern because they are better able to describe the rebound of choices of advised options after the exploration phase.

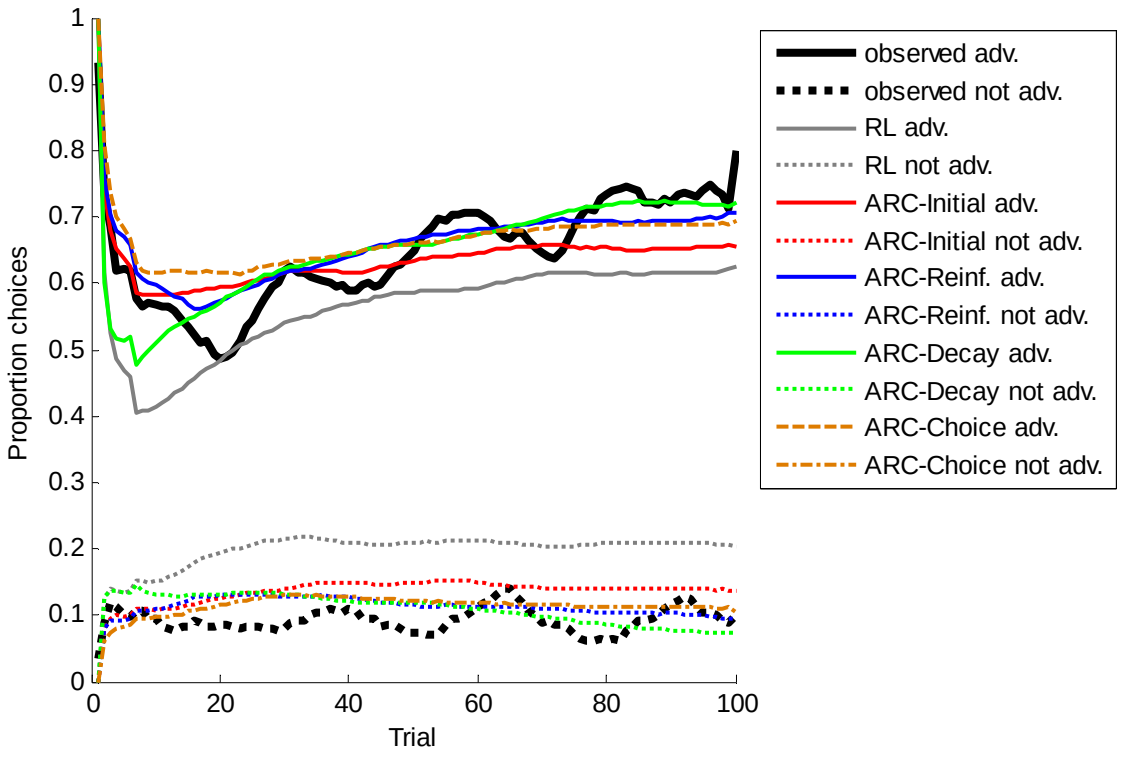

Figure 4.5. Proportion of participants choosing the advised deck (adv.) or the corresponding deck with the same average payoff (not adv.). Comparison of observed and simulated choices.

4.5 Discussion

The aim of Experiment 4.1 was to examine if people use advice, if advice improves performance, and which of the ARC models describe the learning process. The first question has to be answered positively because advice receivers generally preferred the decks that were advised to them. However, only few (5) receivers followed the advice for all 100 choices of the IGT. The majority of receivers started with the advised option, then explored other options, and finally returned to the advised option.

The second question—does advice improve performance?—deserves a conditional answer. Advice receivers performed better than independents, who made decisions without receiving or giving advice, but better than advisors who still had to give advice. Compared to these groups, receivers are approximately 10 percentage points more likely to choose a good deck. However,

receivers perform worse than participants with their own experience in the same task (78% vs. 73% choose a good deck). In sum, receivers have an advantage over inexperienced decision makers.

A finding needing explanation is that advisors do not start with the deck they gave as advice when they made their second 100 trials of the IGT. I can only speculate that advisors either, first, did not believe that they chose from the same decks again, or, second, that they had learned that one could choose from the two bad decks at least in the first trials.

The third question—which model describes social learning best?—receives a first, but not final answer. Examining ΔAICs, I found—in agreement to the answer to question 1—that models implementing social learning perform better than those without. Comparing the social learning models, I found that ARC-Initial is, on average, least able to describe participants' choices. This is reflected in the worse fit, compared to ARC-Reinforcement, ARC-Decay, and ARC-Choice and, more importantly, in the weaker ability of ARC-Initial to account for the characteristic adherence-exploration-adherence choice pattern of most participants. While it seems clear that RL and ARC-Initial cannot explain the choices of the majority of participants in a satisfactory manner, a decision between ARC-Reinforcement, ARC-Decay, and ARC-Choice is more difficult. Even though ARC-Reinforcement has a better fit than its remaining competitors, the difference is small, compared to the differences to other models. Additionally, the simulation in Experiment 4.1 showed that ARC-Decay and ARC-Reinforcement (and partly also ARC-Choice) explain receivers' adherence to advice similarly well.

An alternative analysis examined, for how many participants each model was the best model. This analysis is justified because, already, the raw data revealed qualitatively different learning types. One type of participants describes a learning process that includes no individual learning, but simply takes the advice and uses it for all choices—describing 5 participants—and another, more frequent, type of participants combines both information: advice and individual performance feedback. A clear dominance of ARC-Reinforcement over the other models seems questioned if one considers how often a model is the best model for the participants. ARC-Decay performs best in this regard, and ARC-Reinforcement is second-best. In sum, the model comparison showed that, overall, ARC-Reinforcement describes participants' choices best, closely followed by ARC-Decay, and partly also ARC-Choice. At the same time, I find that different people seem to learn differently.

In sum, Experiment 4.1 showed that decision makers who receive advice use it to perform better than inexperienced decision makers. I found that social learning models describe individuals' choices better than the individual reinforcement learning model. However, ARC-

Reinforcement, ARC-Decay, and partly also ARC-Choice perform similarly well, hence, Experiment 4.2 will further examine which of the three models are more appropriate to describe social learning. A peculiarity of Experiment 4.1 was that participants rarely received bad advice, therefore, in Experiment 4.2, I will examine how bad advice influences learning, and if the social learning models can describe learning after bad advice.

4.6 Experiment 4.2

This experiment compares ARC-Reinforcement, ARC-Decay, and ARC-Choice in two ways. First, I will test the three models' predictions based on the estimated parameters of Experiment 4.1, thus, I will not fit the models to the data from Experiment 4.2. This provides a strong generalization test of the models. In Experiment 4.1, I compared the models according to the ΔAIC criterion. This approach has been criticized because the AIC does not take the models' functional form into account, so that it might not adequately reflect the complexity of a model (Myung & Pitt, 1997). For this reason, alternative model comparison techniques have been proposed (Pitt, Myung, & Zhang, 2002). I will follow the generalized criterion methodology (Busemeyer & Wang, 2000). Here, models are evaluated by using one experiment to estimate the models' parameters, and thereafter, on the basis of the estimated parameters predictions for a new situation, are determined to perform the model comparison test. Thus, Experiment 4.2 represents the crucial generalization test of ARC-Reinforcement and ARC-Decay.

For an illuminative test, it is further desirable to find a situation in which the models make different qualitative predictions. For the qualitative prediction, I focused on the two best models in Experiment 4.1, ARC-Reinforcement and ARC-Decay. These two models make qualitatively different predictions when individuals encounter a multi-armed bandit task in which all decks have negative expected payoffs. ARC-Reinforcement predicts that participants still prefer the advised deck when expected payoffs are in the loss domain. Specifically, ARC-Reinforcement predicts that the advised deck should be selected with the highest probability, even when the advised deck has a lower expected payoff than alternative options, as long as the payoff difference to the better deck is smaller than the additional reinforcement for the advised deck. ARC-Decay predicts the opposite, that is, in a situation in which the options have negative expected payoffs, individuals should avoid the advised option. The reason is that, due to the smaller decay rates for the advised decks, ARC-Decay predicts a long memory for negative payoffs, that is, propensities remain for a longer time negative longer after losses. Compared to the advised option, the other options will appear more attractive because the higher decay rate

implies a short memory for the negative payoffs, so that propensities, and thus choice probabilities, will be higher than for the advised decks.

ARC-Reinforcement and ARC-Choice both predict that advice receivers should prefer the advised deck over the corresponding deck. However, only ARC-Choice explicitly predicts stronger adherence to advice when advice receivers are more uncertain about which options are better. This uncertainty increases as a function of the difference between the payoffs from the options and as a function of the standard deviation of options' payoffs. Uncertainty can be expressed as the effect size of the payoff difference between good and bad decks. The effect size of the original IGT is $D = 0.24$, calculated as the difference in payoffs from good and bad decks divided by the standard deviation of the pooled payoffs from all options, so that a smaller effect size indicates higher uncertainty.

Accordingly, a payoff schedule that allows distinguishing between the three remaining models has predominantly negative payoffs, allowing a distinction between ARC-Decay and ARC-Reinforcement, and a small effect size for the payoff difference between good and bad decks, allowing a distinction between ARC-Choice and ARC-Reinforcement. In line with these demands I devised a payoff schedule where payoffs are drawn form a normal distribution with a standard deviation of 20 euro-cents, where the good decks have an expected payoff of -7 euro-cents, and the bad decks of -10 euro-cents. The effect size for this payoff schedule is $D = 0.15$.

The computational models allow us to make quantitative predictions for advice receivers' choices in the new task. To determine the models' predictions, I applied a nonparametric bootstrapping procedure to select model parameters from the first experiment. Specifically, a participant in Experiment 4.2 was simulated by using the parameters of one participant in Experiment 4.1, simulating the task 100 times, and calculating the expected choice probabilities of the simulated participant as the mean over the 100 simulations. As the influence of social learning in ARC-Initial and ARC-Reinforcement is defined as a function of the expected payoff of the best option in the set, the parameters in Experiment 4.1 can be applied without scaling them. To obtain average choice probabilities for one "virtual experiment," the parameters of 30 randomly selected participants (with replacement) of Experiment 4.1 were matched with random advices, which were constrained so that 50% of the advices was good and 50% bad. The predictions of the models in Experiment 4.2 were determined by averaging over 5,000 virtual experiments. Figure 4.6 displays the predictions of the three models for adherence to advice in Experiment 4.2.

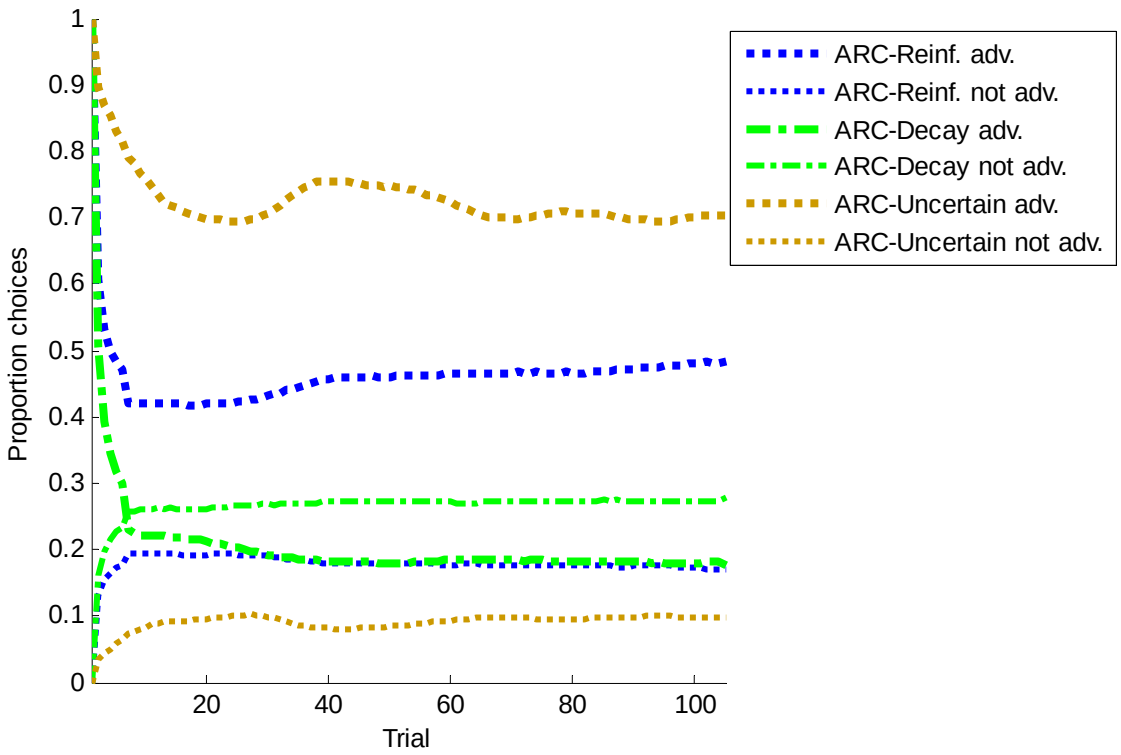

Figure 4.6. Predictions of the social learning models for the task in Experiment 2. ARC-Choice predicts high adherence rates, ARC-Reinforcement predicts lower adherence rates, and only ARC-Decay predicts that the advised deck will be chosen less frequently than the corresponding deck.

4.6.1 Method

4.6.1.1 Design

Experiment 4.2 used a multi-armed bandit task similar to the IGT in Experiment 4.1. The key difference to Experiment 4.1 was the payoff schedule. Deviating from the original payoff schedule from Bechara et al. (1994), average payoffs in Experiment 4.2 were –10 euro-cents for the bad decks and –7 euro-cents for the good decks, all decks had a standard deviation of 20 cents. Payoffs were randomly drawn from a normal distribution, and mean and standard deviation of payoffs were maintained in blocks of 15 trials; I label this the standard payoff schedule. Participants started with an endowment of 12.5 euro and made 105 choices. Figure 4.6 displays the predictions of the three models for adherence to advice in Experiment 4.2, assuming that 15 receivers had good and 15 receivers had bad advice.

One result from Experiment 4.1 was that participants rarely received bad advice. Pilot tests for Experiment 4.2 revealed that it was difficult to find a payoff schedule for which approximately half of the participants learn which are the better decks and, hence, give good

advice. Therefore, I made the task for 20 of the 30 advisors in Experiment 4.2 more difficult by manipulating their payoff schedule (henceforth nonstandard advisors), so that approximately half of the participants would receive good, respectively bad, advice. The manipulation consisted of subtracting 5 euro-cents from every payoff of a good deck in the first 30 trials, and adding 5 euro-cents to 30 randomly selected trials out of the last 75 trials. As in Experiment 4.1, the task was performed by advisors, receivers, and independents, with the latter two always choosing from the standard payoff schedule.

4.6.1.2 Participants and Procedure

Eighty participants, mostly students from the Free University Berlin (55% women, with a mean age of 25 years), were randomly assigned to the three conditions. Thirty participants were advisors, 30 were receivers, and 20 independents. With two exceptions, the experimental procedure was identical to Experiment 4.1. 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). Second, participants' variable payoff was calculated without subtracting the initial endowment because the expected payoff of all decks was negative.

4.6.2 Results of Experiment 4.2

Choices and performance. Participants earned, on average, 3.83 euro (*SD* = 1.9) in the IGT. Figure 4.7 depicts the proportion of participants who chose one of the two good decks, and shows that the task in Experiment 4.2 was more difficult for most participants because the average probability to choose a good deck was, compared to Experiment 4.1, lower throughout the task. Ten Standard advisors chose a good deck in, on average, 60% of all trials (*SD* = 17%) in their first and 69% (*SD* = 10%) in their second 100 trials. Twenty non-standard advisors chose a good deck in, on average, 39% of all trials (*SD* = 12%) in their first and 42% (*SD* = 40%) in their second 100 trials. The worse performance of nonstandard advisors was expected because their task was more difficult. Thirteen receivers with good advice chose a good deck, on average, in 69% (*SD* = 15) of their trials, 17 receivers with bad advice in 48% (*SD* = 20), and independents in 63% (*SD* = 12). Comparing advisors, receivers, and independents, I found that receivers with good advice performed better than those with 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 with bad advice performed worse than independent learners, $t(35) = 2.73$, $p = .01$, $d = .9$.

To examine if independent participants learned to choose the good decks, I tested if they chose a good deck in more than 50% of all trials. The significant result indicates that

independent players, on average, learned to choose a good deck, $t(19) = 4.71$, $p < .001$, $d = 1.05$. In 70% (SD = 9%) of the last 50 trials they chose a good deck, 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 4.2, the receivers benefit less from the advice, compared to Experiment 4.1. Nevertheless, the receivers still profited from receiving good advice, but were harmed by bad advice. Independents still learned to choose the good decks.

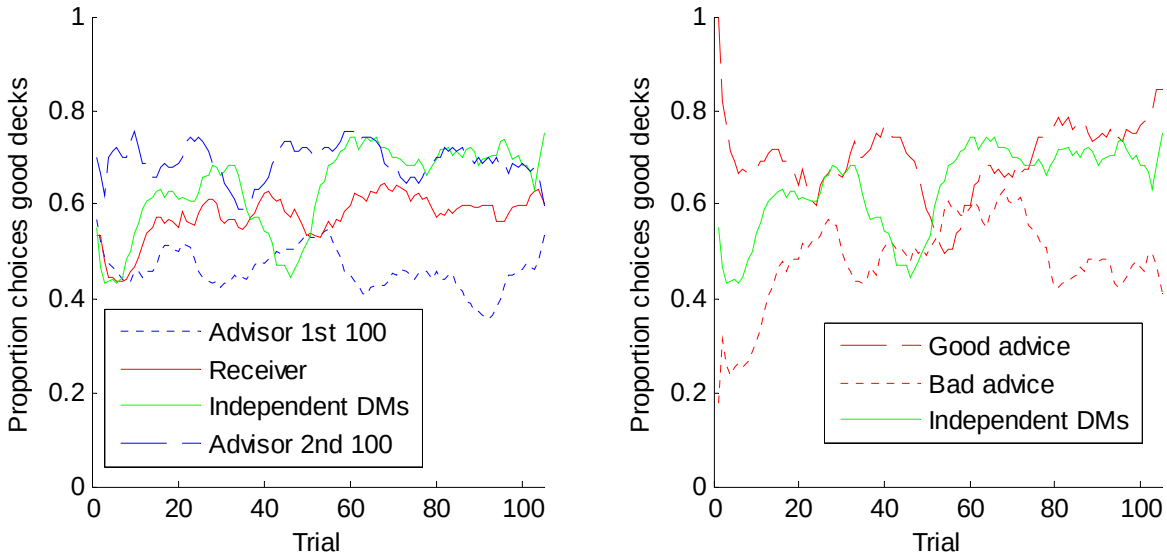

Figure 4.7. Proportion of participants who chose a good deck distinguished for independents, advisors, and receivers.

Advice giving and following. Three out of 10 standard advisors and 14 out of 20 nonstandard advisors advised to choose a bad deck. This result was expected because the payoff schedule for nonstandard advisors made their task more difficult. The relative frequencies of choices of advised decks and corresponding decks with the same expected payoff show—equivalent to Experiment 4.1—if advice influenced choices. Out of 30 receivers, .56% started the IGT by choosing the advised deck. Figure 4.8 displays the proportion of receivers who chose the advised or the corresponding deck (see also Table 4.4). To test the influence of the quality of advice, I performed a repeated measurement, ANOVA, with the quality of advice (good vs. bad) as a between-subject factor, advice (advised deck vs. corresponding deck) as a within-subject factor, and the choice frequency of decks as a dependent variable. The main effect for the quality of advice was significant, $F(1,28) = 6.38$, $p = .017$, $\eta 2 = .19$, as was the main effect for advice, $F(1,28) = 47.51$, $p < .001$, $\eta 2 = .063$). The interaction of advice and quality, and advice, approached the conventional significance level, $F(1,28) = 3.78$, $p = .063$, $\eta 2 = .12$. These results

show that receivers adhered more to advice when this was good, and that the probability to choose the corresponding deck was not influenced by the quality of advice. The larger effect size for the factor advice, compared to the factor quality of advice, indicates that participants' choices were more influenced by advice than by the payoffs from choosing the decks.

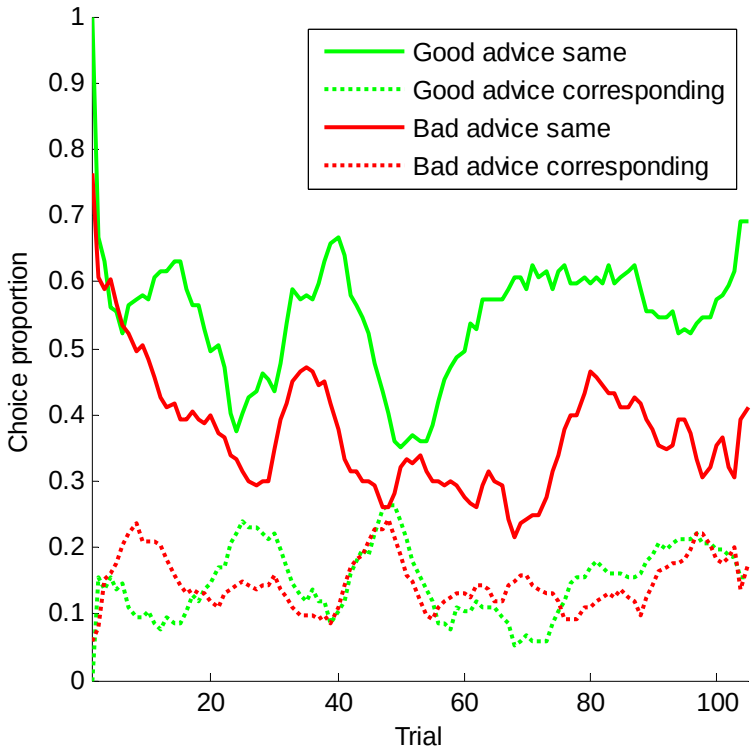

Figure 4.8. Proportion of participants choosing the advised decks (indicated by "same") or the corresponding decks with the same expected payoffs (moving average, with window size 11).

Model comparison. A first step to compare the models is to examine their qualitative predictions. According to ARC-Reinforcement, participants who received bad advice should adhere to the bad deck in the presence of better alternative decks more often than participants who received no advice. This can also be assumed for ARC-Choice because the variance of payoffs from decks was high, compared to expected payoffs. In contrast, ARC-Decay predicts that receiving bad advice should decrease the probability that receivers choose the advised bad deck. Differently than predicted by ARC-Decay, Figure 4.8 shows that players who received bad advice choose the good decks less frequently than independent decision makers. Further, Figure 4.8 shows that—as predicted by ARC-Reinforcement and ARC-Choice—receivers chose the nonadvised deck less frequently than the advised deck, and not the opposite as predicted by ARC-Decay.

Additionally, to compare the models, I examined how well they predict participants' choices. To compare the models' predictive power, I performed a generalization test (Busemeyer & Wang, 2000), that is, I determined the models' predictions based on the estimated parameters in Experiment 4.1. The predictions of the models were simulated as described above, except that parameter values from participants in Experiment 4.1 were randomly matched with real advices from Experiment 4.2. Specifically, I examine how good the models predict choices of the advised deck and the corresponding deck with the same expected payoff. Figure 4.9 shows the observed and simulated choice proportions, and shows that ARC-Reinforcement predicts the preference for the advised decks better than ARC-Decay and also better than ARC-Choice, which clearly overestimates the influence of social learning in Experiment 4.2. This result is also supported by the RMSDs illustrated in Table 4.4, which show that ARC-Reinforcement predicts the likelihood to choose the advised deck better than any other model. Table 4.4 also shows that ARC-Reinforcement does not only predict mean choice proportion better than ARC-Reinforcement but also the standard deviations observed in Experiment 4.2. Thus, the generalization test supports ARC-Reinforcement as the best considered learning model.

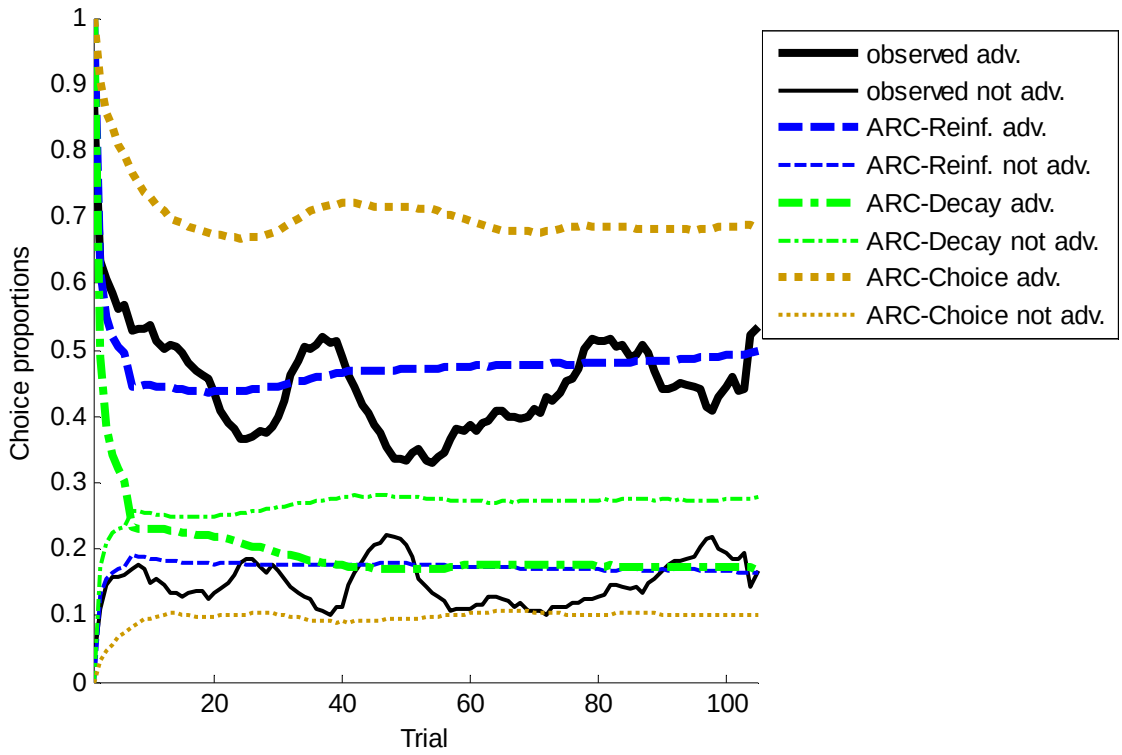

Figure 4.9. Observed and predicted probability to choose the advised deck or the corresponding deck with the same expected payoff.

Table 4.4. *Results of the bootstrapping simulation.*

Advice	Deck	Data	RL	ARC			
				Initial	Rein-forcement	Decay	Choice
Good	Good	.7 (.15)	.56 (.04)	.63 (.15)	.68 (.16)	.49 (.06)	.83 (.18)
	Advised	.55 (.18)	.29 (.02)	.42 (.23)	.5 (.25)	.22 (.07)	.72 (.3)
	Corresponding	.15 (.13)	.27 (.02)	.21 (.09)	.18 (.09)	.28 (.06)	.1 (.11)
	RMSD						
	Good deck	-	.14	.088	.06	.205	.127
	Adherence	-	.207	.113	.069	.256	.129
	All choices	-	.11	.081	.068	.137	.086
Bad	Good	.48 (.2)	.55 (.04)	.46 (.13)	.37 (.18)	.56 (.07)	.22 (.23)
	Advised	.37 (.19)	.22 (.02)	.35 (.18)	.46 (.26)	.19 (.08)	.69 (.32)
	Corresponding	.15 (.08)	.22 (.02)	.19 (.06)	.17 (.08)	.25 (.03)	.09 (.09)
	RMSD						
	Good deck	-	.094	.077	.12	.084	.263
	Adherence	-	.122	.067	.08	.154	.231
	All choices	-	.102	.075	.1	.109	.157
Good and bad	Good	.58 (.21)	.55 (.04)	.53 (.16)	.5 (.23)	.53 (.07)	.48 (.36)
	Advised	.45 (.2)	.26 (.03)	.38 (.21)	.48 (.26)	.2 (.08)	.7 (.31)
	Corresponding	. 15 (.1)	.24 (.03)	.2 (.08)	.17 (.08)	.26 (.04)	.1 (.1)
	RMSD						
	Good deck	-	.041	.046	.073	.049	.095
	Adherence	-	.155	.078	.051	.197	.184
	All choices	-	.081	.065	.068	.089	.075

Note. Results of the bootstrapping simulation: Probability to choose good decks, advice following advice for good and bad advice, and RMSDs. Cells depict means (SD), predicted standard deviations were calculated as the mean of the standard deviations over all simulated experiments.

Do the models predict receivers' behavior independently of the quality of advice? To examine this question, I performed the same bootstrapping procedure as above, separately for receivers of good and bad advice. Table 4.4 displays the results of these simulations and shows that when advice is good, ARC-Reinforcement predicts adherence to advice well, whereas ARC-Decay underestimates adherence as well as the portion of choices of good decks, and ARC-Choice overestimates both. When advice is bad, ARC-Decay again underestimates adherence to advice and chooses again—as predicted—the nonadvised deck more frequently than the advised deck. ARC-Choice clearly overestimates adherence to advice and, hence, predicts low proportions of choices of the good decks. To a lesser degree this is also true for ARC-Reinforcement, which, nevertheless, still correctly predicts stronger adherence to advice when advice is good, 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 for bad advice also indicates that ARC-Reinforcement is the best model.

4.6.3 Discussion of Experiment 4.2

As intended, the second experiment allowed us to examine social learning in a situation in which the best social learning models in Experiment 4.1 made rather different predictions and under conditions of bad advice. I found again that receivers generally use advice. In addition to the results in Experiment 4.1, I found that advice had a greater impact on receivers' choices, compared to the payoffs they received from choosing the options. Regarding the effect of advice, good advice improved performance and bad advice harmed performance.

Experiment 4.1 showed that advisors did not choose the decks that they had advised when they started with their second 100 trials. In contrasts, advisors in Experiments choose the deck they had advised from the beginning on in their second 100 trials. This suggests that participants did not distrust the instructions, and that advisors in Experiment 4.1 had, indeed, learned that one can choose from the bad decks in the first trials without risking high losses.

Experiment 4.2 compared ARC-Reinforcement, ARC-Decay, and ARC-Choice by means of qualitative and quantitative predictions. The qualitative prediction of ARC-Decay has to be rejected because participants receiving bad advice chose bad decks more frequently than independent participants. Further, receivers chose the advised deck more frequently than the corresponding deck with the same expected payoff. These results are only in line with the prediction from ARC-Reinforcement and ARC-Choice. The simulated predictions of the models in Experiment 4.2 also favor ARC-Reinforcement because ARC-Choice predicted a too high adherence to advice. Only ARC-Reinforcement correctly predicts receivers' adherence to advice,

and predicts more adherence to good than to bad advice, even though the predicted difference is smaller than the observed difference. Finally, ARC-Reinforcement also predicted the experimental variance.

Experiment 4.1 showed that different participants are best modeled by different learning models. As the models were not fitted to participants in Experiment 4.2, it is not possible to classify participants to models. I nevertheless argue that ARC-Reinforcement is not only best, on average, but also best describes most participants. The reason is the finding that the proportion of participants with bad advice who chose a bad deck is stable throughout the last 70 trials and consistent with ARC-Reinforcement, whereas ARC-Initial and ARC-Decay predict constantly decreasing choice frequencies. Interestingly, contrary to Experiment 4.1, no participant in Experiment 4.2 followed the advice in all 100 trials. While this can be a coincidence, one might speculate that participants in Experiment 4.2 experienced losses in most trials, which might stimulate a stronger exploration of alternative options.

In sum, Experiment 4.2 clearly supports ARC-Reinforcement as the best model to describe the participants' social learning process because it predicts adherence to advice, conditional on the quality of advice, and also the variance of Experiment 4.2.

4.7 General Discussion

The aim of this paper was to examine social learning, when people make repeated choices from experience, by answering three questions. Do people use advice? Does advice-taking improve decision performance? How can social learning be best described? To examine these questions, I observed choices in two multi-armed bandit tasks with four choice options, and tested one model of individual and three models of social learning.

The first result was that advice receivers used advice because they chose the advised deck more frequently than other decks, in particular the corresponding deck with the same average payoff. Moreover, in Experiment 4.2, advice receivers even followed the bad advice of choosing decks with the lowest expected payoff. The influence of advice was also visible in receivers' performance; as in Experiment 4.1, receivers with good advice performed better than independent decision makers, who were, in turn, better than participants receiving bad advice. The bad performance of receivers who received bad advice in Experiment 4.2 shows that social influence can distract people from learning independently. That participants adhered more to good advice than to bad advice shows that they, indeed, combined individual experience and social information to make choices.

To predict how people combine advice and their own experience, I proposed and tested four models of social learning, and compared them with an individual learning model. All models are modifications of existing reinforcement learning models (Erev, 1998; Yechiam & Busemeyer, 2005). ARC-Initial assumes that the advised options are initially evaluated more positively, compared to alternative options, expressed in the model by higher initial propensities. ARC-Reinforcement assumes that payoffs from advised options lead to stronger reinforcements. ARC-Decay assumes that propensities of advised options decay slower. Finally, ARC-Choice assumes that people choose the advised options when the propensities of options are similar. A first finding confirming my general modeling approach is that, in Experiment 4.1, all social learning models described choices better (in terms of ΔAICs and RMSDs), compared to the statistical baseline model and the individual reinforcement learning model. A comparison of the social learning models identified ARC-Reinforcement, ARC-Decay, and ARC-Choice as the best social learning models because these models better predicted participants' choice frequencies of good decks, the level of adherence to advice, and the typical adherence-exploration-adherence choice pattern.

Experiment 4.2 tested among ARC-Reinforcement, ARC-Decay, and ARC-Choice by implementing another version of the multi-armed bandit task, similar to the IGT, in which payoffs for all decks were negative. In such a situation, the two best models identified in Experiment 4.1 make diverging predictions. Contrary to the prediction of ARC-Decay, and in agreement with the prediction from ARC-Reinforcement and ARC-Choice, Experiment 4.2 showed that participants kept choosing the advised option, which, in the case of bad advice, implied that the learning process did not allow them to find the best option to choose from. Whereas ARC-Choice generally overestimated adherence to advice, ARC-Reinforcement correctly predicted that adherence to advice is higher when advice is good, and also predicted the variance of choice proportions in Experiment 4.2. In sum, the experiments show that decision makers adhere to advice, that good advice helps and bad advice harms learning, and that ARC-Reinforcement describes the social learning process best.

While I implemented the different assumptions of social learning separately, one might argue that individuals are better described by implementing (some of) these into the same model. I refrained from this possibility because this would increase the complexity of the social learning model, and make it more difficult to understand how the mechanisms change the learning process, compared to individual learning. Further, ARC-Reinforcement alone could already describe the data well. Nevertheless, future research will have to examine combinations of the models.

When social learning is truly different from individual learning, social information should influence individual learning differently than one's own experience. That ARC-Initial did not adequately model the learning process indicates that the learning process of advice receivers cannot be modeled by assuming that advice only influences the initial preference of decision makers, 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 than one's own experience is also supported by the finding that advisors in their second 100 trials behaved differently than advice receivers. Figures 2 and 7 show that advice receivers explore alternative options longer than advisors in their second 100 choices. This is particularly clear in Experiment 4.2, where advisors in their second 100 trials choose the good decks with nearly constant probability over all choices.

How does social learning differ from individual learning? When comparing the social learning models, I highlighted the point in time at which social information is effective, that is, the persistence of social information and the effect of bad advice on learning. The results show that social information influenced learning in the first choices and especially after the exploration phase. This is mostly consistent with the characterization of ARC-Reinforcement, for which the influence of social learning increases as the additional reinforcement from choosing the advised option accumulates. Accordingly, participants' behavior is also consistent with the second characterization of the ARC-Reinforcement, that the influence of social information is persistent over time. Importantly, bad advice leads to worse performance compared to choosing without advice. In sum, social learning as identified in the experiments, and modeled by ARC-Reinforcement, is characterized by the combination of social information and individual experience. Social information determines choices briefly at the beginning and stronger and persistently after the exploration phase.

As the results in Experiment 4.2 show, sensitivity to 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. While the experiment was arranged so that participants had every reason to assume that advisors were competent, Yaniv and Kleinberger (2000) explicitly examined how decision makers assess advisors' reputation and found that decision makers update appraisals asymmetrically. Specifically, a good appraisal was quickly downgraded after bad advice—that is, participants lowered the weight of social information, compared to their own information—and an appraisal was only slowly upgraded after good advice. In a further experiment, Yaniv and Kleinberger found that decision makers can judge advisors' performance

accurately, and purchase information only from good advisors. Luan et al. (2004) also found that advice receivers can correctly judge the performance of different sources of advice. These results, while obtained in different experimental settings than the experiments reported in this article, indicate that advice receivers possess mechanisms allowing them to select good advisors, thus hedging the sensitivity of their social learning mechanism to bad advice. Beyond this argumentation, experiments by Celen, Kariv and Schotter (2005), Kameda and Nakanishi (2003), and Yaniv (Yaniv, 2004a), show that even naïve participants tend to give useful advice, and that social learning generally improves performance.

While the models describe the social learning process on the computational level, some questions remain open due to limitations in the experiments. One unresolved issue is, if all decision makers can be described with the same model, or if people have qualitatively different learning processes. Experiment 4.1 suggested that different participants are best described with different models, and the analysis of Experiment 4.2 suggested than one model is sufficient to describe all receivers. While the results about the heterogeneity of participants, beyond different model parameters, are not conclusive, the good performance of ARC-Reinforcement in predicting mean and variance in Experiment 4.2 suggests that a single model can predict most individuals' behavior.

A second limitation concerns the psychological processes driving the ARC-Reinforcement model. I motivated ARC-Reinforcement with the intuitions that payoffs from the advised option lead to higher reinforcement, due to a confirmation bias in the evaluation of information or due to the intrinsic value of the adherence to advice following. While the experiments could not examine these assumptions, related research on individual decision making supports at least the confirmation bias interpretation. Betsch et al. (2001) found that people who had learned to choose an option in a repeated decision-making task suppressed disconfirming information and overvalued confirming information when the decision environment changed to favor another option. While this result was obtained after the preference for one option was induced by the own experience, I assume that the same process can be triggered when the preference is induced by social information. A final test between the confirmation bias hypothesis and the reward from conformity hypothesis could exploit that the two approaches make different predictions about the influence of the source of advice. While the confirmation bias hypothesis predicts that the source of advice plays no role, the reward from conformity hypothesis predicts stronger adherence to advice from humans than, for instance, from computers.

A third limitation concerns ARC-Reinforcement's overestimation of adherence to advice after bad advice. While this suggests that the individual learning component of participants'

behavior in Experiment 4.2 was underestimated, this is probably not due to the general inability of ARC-Reinforcement to explain participants' behavior. The reason rather lies in the fact that all participants in Experiment 4.1 received good advice, so that individual and social learning usually pointed in the same direction. In this case, it is likely that one selects relatively high social learning parameters because it predicts similar behavior expected from individual learning. The high social learning parameters might then have suppressed ARC-Reinforcement's individual learning component in the simulations used to predict behavior in Experiment 4.2.

The situation in which I examined social learning is specific in that players were always only informed about the payoffs from chosen options, in that participants received social information only once, and in that social information was given as explicit advice instead of providing the opportunity to observe others' choices. I argue that the ARC models, especially ARC-Reinforcement, can be applied to different situations with minor modifications. First, when participants are informed about forgone payoffs, ARC-Reinforcement could simply be modified to allow for the updating of nonchosen alternatives, for instance, as described by Camerer and Ho (1999b). Specifically, when forgone payoffs are also used to update propensities, the portion of social reinforcement on the sum of reinforcements will be smaller, which implies less influence of social information. ARC-Reinforcement could also be used to model continuous advice in every trial, by adding a constant to payoffs when reinforcing the currently advised options, just as I did for a single peace 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 proportion of others who were observed choosing these options (for an implementation of such a model see McElreath et al., 2004). When decision makers are informed about others' choices and payoffs, choice options could be reinforced by their own and by the others' (discounted) payoff. While the application of ARC-Reinforcement to different types of social learning is straightforward, experimental investigation will have to verify how the performance of ARC-Reinforcement generalizes to different social learning situations.

To examine the mechanism of social learning in repeated choice tasks, or decisions from experience, I used the abstract paradigm of the multi-armed bandits. As this paradigm reflects important characteristics of many choice domains, such as decisions from experience (Barron & Erev, 2003; Erev & Barron, 2005; Hertwig et al., 2004), consumer choice (e.g. Anderson, 2000; Erdem & Keane, 1996; Meyer & Shi, 1995), choice among decision strategies (e.g. Rieskamp & Otto, 2004; Siegler & Araya, 2005), and interdependent decision making (e.g. Hanaki et al., 2005; Stahl & Haruvy, 2002), I expect that ARC-Reinforcement can be used to model social

learning in these domains. However, future research will need to examine how well ARC-Reinforcement actually predicts behavior in these situations, which are all more complex than the abstract multi-armed bandit paradigm.

4.8 Conclusion

The aim of this research was to investigate how advice influences learning in repeated choice tasks. I found that people use advice, so that the influence of social information is strong for the first choices, is reduced in the following exploration phase, and persistently gains influence when the task continues. Advice receivers are more likely to choose the advised option when it is the best option in the set, compared to when better alternatives are available. The model that best accounted for this learning pattern was ARC-Reinforcement, a social learning model that inputs social information in a simple reinforcement learning process by assuming that payoffs from advised options lead to higher reinforcements.