Cognitive Psychology 78 (2015) 78–98

Contents lists available at ScienceDirect

Cognitive Psychology

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

Of matchers and maximizers: How competition shapes choice under risk and uncertainty

Christin Schulze [a,*], Don van Ravenzwaaij [a,b], Ben R. Newell [a]

[a] School of Psychology, University of New South Wales, Sydney, Australia
[b] School of Psychology, University of Newcastle, Newcastle, Australia

ARTICLE INFO

Article history:
Accepted 17 March 2015
Available online 8 April 2015

Keywords:
Probability matching
Reinforcement learning
Competition
Cognitive models
Decision making

ABSTRACT

In a world of limited resources, scarcity and rivalry are central challenges for decision makers—animals foraging for food, corporations seeking maximal profits, and athletes training to win, all strive against others competing for the same goals. In this article, we establish the role of competitive pressures for the facilitation of optimal decision making in simple sequential binary choice tasks. In two experiments, competition was introduced with a computerized opponent whose choice behavior reinforced one of two strategies: If the opponent probabilistically imitated participant choices, probability *matching* was optimal; if the opponent was indifferent, probability *maximizing* was optimal. We observed accurate asymptotic strategy use in both conditions irrespective of the provision of outcome probabilities, suggesting that participants were sensitive to the differences in opponent behavior. An analysis of reinforcement learning models established that computational conceptualizations of opponent behavior are critical to account for the observed divergence in strategy adoption. Our results provide a novel appraisal of probability matching and show how this individually 'irrational' choice phenomenon can be socially adaptive under competition.

© 2015 Elsevier Inc. All rights reserved.

* Corresponding author. Address: Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Fax: +49 30 82406490.
E-mail address: cschulze@mpib-berlin.mpg.de (C. Schulze).

http://dx.doi.org/10.1016/j.cogpsych.2015.03.002
0010-0285/© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Competition is a pervasive characteristic of the world—plants compete for light, water and pollination; animals are in continual competition for food, territory and procreation; and humans constantly compete in sports, for social status and companionship. The presence of social competitors in virtually all aspects of real-life decision making demands the development of socially adaptive choice strategies in a broad range of contexts. In this article, we examine how competitive pressures shape the adequacy of decisions made in simple binary choice probability learning paradigms (cf. Estes, 1964) that to our knowledge have only been employed with individual decision makers in social isolation. This research closes an important conceptual gap in the human probability learning literature by offering a socially adaptive account for a long-standing individual choice phenomenon known as *probability matching*.

Probability matching describes an inferior strategy of sampling choice options in proportion to the options' relative outcome probabilities in sequential probability learning tasks. The standard procedure in such choice tasks involves asking individual decision makers to choose repeatedly between two alternatives that reward the same payoff with unequal odds (e.g., with $p = .70$ and $1 - p = .30$). Assuming these outcome probabilities are stationary, payoffs are maximized by exclusively selecting the option with the higher reward likelihood—i.e., by *probability maximizing*—once the probabilities have been learned. By contrast, matching choices to outcome frequencies by choosing the more likely option on 70% of occasions and its alternative on the remaining 30%, results in markedly inferior choice accuracies and payoffs. Yet, probability matching is commonly adopted by individual decision makers (for a review see Vulkan, 2000). Context-independent interpretations of rational choice therefore consider probability matching as an erroneous choice anomaly and attribute its adoption to cognitive constraints of the decision maker. Within the framework of dual cognitive process theories, for instance, probability matching is assumed to represent a simple cognitive shortcut carried out by an intuitive cognitive system, whereas probability maximizing would arise if deliberation corrected this initial impulse (Koehler & James, 2009, 2014; Kogler & Kühberger, 2007; West & Stanovich, 2003).

What seems irrational in individualized context-free environments, however, can be optimal in ecologically plausible situations. Accordingly, an alternative view on probability matching holds that this tendency may emerge as a result of over-generalizing typically highly adaptive behaviors, for example, responding to limited and uncertain information about the true random nature of a choice task (Green, Benson, Kersten, & Schrater, 2010), searching for patterns when none exist (Gaissmaier & Schooler, 2008; Peterson & Ulehla, 1965; Wolford, Newman, Miller, & Wig, 2004), or when prospective competitive interactions are taken into account (Gallistel, 1990; Gigerenzer, 2000).

1.1. Probability matching in competitive environments

When decision makers seek to exploit limited resources under *natural circumstances* (e.g., forage for food or make money), they are rarely alone but typically in fierce competition for the exploitation of these resources with other agents. The more individual agents then choose the seemingly richest resource, the less each individual's gain. In nature, this situation cannot remain stable because agents who sometimes select options with potentially scarce resources that are exploitable under less competition would attain a key evolutionary advantage (Gallistel, 1990).

This argument is in line with the predictions of optimal foraging theory (MacArthur & Pianka, 1966), which assumes that animals behave in such a way as to maximize their evolutionary fitness. Within this concept, the model of the ideal free distribution predicts that a group of foragers will distribute their choices among resources relative to the options' reward potential—i.e., probability match—to optimize foraging success (see Fretwell, 1972). This group behavior creates an equilibrated evolutionary stable situation that does not give rise to conditions selecting against it. The predictions of the ideal free distribution have been approximated in various animal studies. Observations of foraging ducks (Harper, 1982), fish (Godin & Keenleyside, 1984), and ants (Lamb & Ollason, 1993) revealed proportionately matched distributions of these animal groups to resource allocations across patches.

Likewise, human group behavior has been assessed in experimental and virtual situations that simulate foraging settings (e.g., Goldstone & Ashpole, 2004; Kraft & Baum, 2001; Madden, Peden, & Yamaguchi, 2002; Sokolowski, Tonneau, & Freixa i Baqué, 1999) and these studies mostly confirm the predictions of the ideal free distribution model. Research on group foraging behavior, however, primarily focusses on the evaluation of group level properties that emerge from collective behavior of individuals but makes little or no predictions regarding individual choices.

Individual choice in situations of conflict, cooperation, and coordination is the main research focus of behavioral game theory (see Camerer, 2003). Research within this framework addresses many key questions also relevant to competitive probability matching. Competitive zero-sum games, for instance, illustrate that people learn to adjust their choices dynamically and approach mixed strategy equilibria successfully (e.g., Rapoport & Budescu, 1992). Evolutionary coordination games—e.g., the hawk-dove game—account for the frequency of aggressive encounters when groups of animals compete for shared resources (Smith & Price, 1973). And results from iterated cooperation games demonstrate that people learn to coordinate and reciprocate in order to maximize efficiency (Axelrod, 1984). Although this research on strategic decision making is extensive, relatively little attention has been paid to exploring the combined requirements of competing against others as well as an uncertain nature, which are characteristic to myriad natural environments (but see e.g., Avrahami, Güth, & Kareev, 2005; Bereby-Meyer & Roth, 2006).

Here, our aim was to examine the role of competition for the facilitation of optimal decision making at the individual choice level in uncertain binary decision contexts. Specifically, the central goal of the present article was to evaluate how ecologically valid competitive pressures shape the adequacy and use of probability matching, as well as its alternative, maximizing. Thus, our approach complements a growing body of literature motivated to reveal the adaptive potential of probability matching under ecologically plausible circumstances (see e.g., Gaissmaier & Schooler, 2008; Green et al., 2010; Wolford et al., 2004) by examining the effect of competition on the phenomenon. Although competitive pressures have been discussed as an underlying cause for probability matching previously (Gallistel, 1990; Gigerenzer, 2000), to date, experimental or analytical demonstrations of optimal human probability matching under competition are lacking in the literature. Our approach adds this key piece of information by assessing human probability matching and maximizing behavior within a competitive choice paradigm.

1.2. A competitive binary choice paradigm

In the tasks employed here, each decision maker competes against a computerized opponent for the exploitation of a monetary resource that an indifferent 'nature' repeatedly places at one of two choice options with unequal probabilities. When both competitors converge on the same choice, potential rewards are split evenly between them. In this paradigm, the success of any strategy depends on the behavior of the opponent. Thus, we manipulated competitor behavior as the key variable. In one condition of our experiments, each computerized opponent's choice probabilities are close imitations of human participant behavior. Considering the prevalence of aggregative behavior in a broad range of natural group settings—e.g., flocking behavior of birds, shoaling of fish, swarming of insects and herd behavior of land animals—a strategy-mirroring opponent creates competitive conditions closely in line with real-life ecological pressures. In a second condition (between-subjects), each human participant is paired with a computer opponent who is indifferent toward her choices. Depending on the encountered competitor type, different choice strategies are optimal.

Under the assumption that the computer competitor is attentive and probabilistically imitates the participant's course of action, probability matching represents the optimal strategy. This is the case because both agents' expected reward proportions are maximized when choice proportions align with outcome probabilities. All studies presented here involved outcome probabilities of $p = .70$ and $1 - p = .30$ and expected reward proportions are defined as the weighted sum of all possible outcomes resulting from nature's move and both agents' choices. When two decision makers follow the same course of action because one imitates the other, the agents' choices can either converge on the same option (requiring them to split potential rewards) or diverge to opposite alternatives (allowing them to keep the full potential payoffs) on any given trial. In the presence of a strategy-mirroring opponent,

the sum of these expected split and full payoff proportions is maximized for each decision maker if both agents probability match, so that $EV_{Matching} = [(.70^3 + .30^3)/2] + [(.70 \cdot .30 \cdot .70) + (.30 \cdot .70 \cdot .30)] = .395$. The left term in brackets denotes the joint probability of both decision makers choosing the same option weighted by the probability of nature rewarding that option and split by two (split payoffs); the right term in brackets expresses the joint probability of the agents selecting different options weighted by the probabilities of nature's move (full payoffs). By comparison, if both decision makers probability maximized, expected rewards would reduce to split payoffs because both agents would exclusively exploit nature's more probable move, so that $EV_{Maximizing} = [(1^2 \cdot .70)/2] = .35$.

Under the assumption that the computer competitor is indifferent toward a participant's choices, probability maximizing represents the optimal strategy.[1] This is the case because sporadic choices by the participant to the inferior option will not tempt an indifferent opponent to do the same and a probability matching opponent can therefore be exploited via maximizing: $EV_{Maximizing} = [(1 \cdot .70 \cdot .70)/2] + [(1 \cdot .30 \cdot .70)] = .445$ (compared to $EV_{Matching} = .395$ for two probability matching agents as above). To facilitate exploration of a range of choice strategies, we provided participants with inter-block feedback that indicated maximum rewards under an optimal strategy. Given the availability of sufficient feedback, we predicted that choices would converge on the respective optimal strategy in both environments as learning progresses (cf. Shanks, Tunney, & McCarthy, 2002): probability matching when competing against a mimicking opponent and probability maximizing when encountering an indifferent opponent.

Learning to choose optimally in our choice paradigm requires a number of cognitive processes also vital for decisions under uncertainty and competition in many real-world situations. These include the exploration of choice profitability and learning about the motives and choice strategies of competing agents. Computational models of reinforcement learning have been shown to successfully and parsimoniously describe such cognitive mechanisms in related tasks (e.g., Busemeyer & Stout, 2002; Erev & Roth, 1998; Gureckis & Love, 2009; Rieskamp & Otto, 2006) and provide an attractive approach to illuminating the nature of learning mechanisms adopted by decision makers in our paradigm. Following the presentation of the behavioral data, we therefore discuss the applicability of a variety of computational models of reinforcement learning that differ with regard to the importance they place on (solely) maximizing profit and out-smarting opponents.

2. Experiment 1

2.1. Method

2.1.1. Participants

Fifty (35 female) undergraduate students from the University of New South Wales with a mean age of 18.92 years (SD = 1.19 years) participated in this experiment in return for course credit and performance based monetary compensation. Earnings ranged from $3.90 to $18.55 (1 AUD ≈ 0.95 USD).

2.1.2. Decision task

A standard probability learning paradigm involving repeated binary decisions with mutually exclusive outcomes over 500 choice trials was used. Choice alternatives were represented by two light bulbs displayed on a computer screen and programmed to illuminate on 70% and 30% of choice trials, respectively—counterbalanced across participants for the left and right choice option. The outcome frequencies were unknown to participants at the start of the experiment and needed to be learned over choice trials. Correct decisions were rewarded with four cents and choices were made while competing against a computerized opponent. When participant and computer opponent converged on the correct response, the payoff was evenly split between them and each received two cents.

[1] This conclusion holds for conditions where one option is at least twice as likely to yield a reward as its alternative—which includes the outcome probabilities considered here (p = .70 and $1 - p$ = .30). For $p \leqslant .66$, payoffs are maximized by exclusively selecting the low probability option if the competitor probability maximizes. We do not consider these cases here.

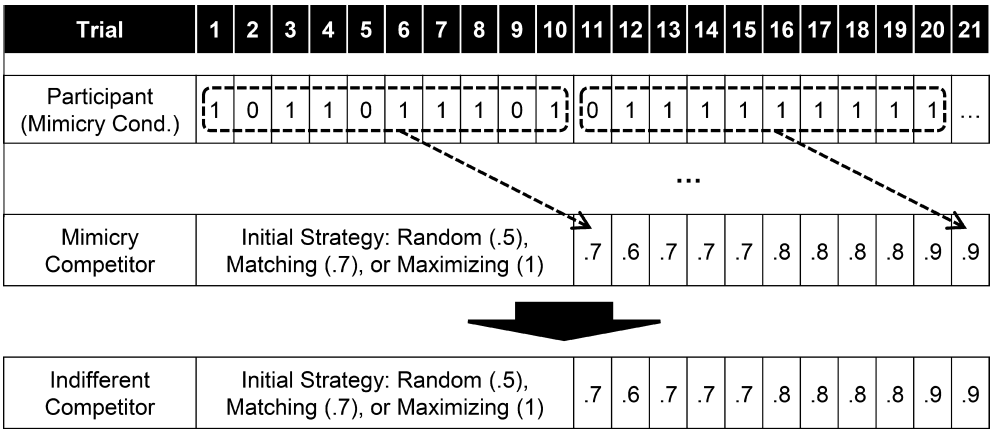

Fig. 1. Computation of opponent behavior for each competitor type condition. Mimicry competitor choice probabilities were based on participant choices over a moving window of the past 10 trials (ones/zeros signify participant choices to the option with the higher/lower outcome probability). Indifferent competitor choice probabilities replayed a probability sequence from the mimicry opponent condition while playing against a participant in the indifferent competitor condition. Therefore, opponent choice proportions were essentially equivalent between conditions but were computed in response to the currently playing participant's behavior (mimicry) or not (indifferent).

2.1.3. Design

We used a 2 × 10 mixed model design with opponent type (mimicry or indifferent) as between-subjects factor and trial block (10 blocks of 50 trials each) as within-subjects factor. Opponent behavior differed between conditions and was computed as illustrated in Fig. 1. During the first 10 trials of the experiment, each competitor in the mimicry condition randomly adopted one of three initial strategies: random response, probability matching, or probability maximizing. Following this period of initial responding, the choice sequence of each mimicry opponent was computed one step ahead by equating the opponent's choice probabilities on a given trial with the choice proportions the participant had displayed during the past 10 trials. For example, if a participant in the mimicry competitor condition chose the more probable option on 7 out of the past 10 trials, her opponent's probability of selecting the same option on the next trial was .70. This algorithm creates opponent behavior that probabilistically mimics participants' choices. By contrast, the choice sequence of each opponent in the indifferent condition was computed irrespective of participants' choices. Instead, the indifferent competitor simply repeated the set of choice probabilities (not the actual choice sequence) that one opponent in the mimicry condition had adopted while playing against another participant. Thus, each opponent in the indifferent condition had a counterpart in the mimicry condition that displayed the same choice *probability* sequence, which renders the two conditions comparable in opponent choice behavior. Using this setup, the behavior of opponents in both conditions may encompass a range of choice strategies (e.g., probability matching, maximizing, or random responding), depending on the choices of participants in the mimicry condition.

Participants were randomly assigned to the mimicry opponent (n = 25, 68% female) or the indifferent opponent type condition (n = 25, 72% female) with the exception of the first/last eight participants who were pre-allocated to the mimicry/indifferent competitor condition to allow adequate computation of yoked opponent behavior as illustrated in Fig. 1. The primary dependent measure was participants' proportion of choices to the more probable choice option.

2.1.4. Procedure

Participants were asked to predict which of two light bulbs would illuminate over a series of 500 trials while attempting to earn as much money as possible. Instructions indicated that the lighting sequence was random in that no pattern or system existed that would make it possible for participants to correctly predict the outcome throughout (see Shanks et al., 2002). Instructions also indicated that

the outcome frequencies for both choice options remained constant during the entire experiment. Additionally, participants in both conditions were informed that a computerized opponent with learning abilities such as their own and no initial information about the lighting frequencies was monitoring their choices and adapting to their skill level. Participants in the mimicry condition also read that the computer opponent would base its decision strategy on their choices. On each trial, participant and opponent simultaneously made a choice, followed by feedback about the other agent's decision and the outcome (one light bulb lit up). Feedback about the participant's and the computer's choices as well as the outcome remained available for a moving window spanning 10 trials back.

Upon completion of every block of 50 trials, a self-paced pause interrupted the experiment, block feedback was provided, and a short message reminded participants that the lighting sequence was random. Participants were told: "In this game you earned $X. Using an optimal strategy you could have earned at least $Y.", where X represented the actual earnings of that block and Y was computed by an optimizing algorithm (Shanks et al., 2002). This algorithm was set to probability matching in the mimicry opponent condition and probability maximizing in the indifferent opponent condition and took both agents' actual decisions during that trial block into account. Additional incentives to improve performance on the following block were provided by informing participants that reaching optimal performance (±three cents) would double their payoff, whereas suboptimal performance would result in halved earnings on the subsequent trial block. Following the decision task and display of total earnings, participants were paid in cash by the experimenter.[2]

2.2. Behavioral results and discussion

In addition to conventional null hypothesis significance testing, we conducted Bayesian analyses based on the default Bayesian ANOVAs outlined by Rouder, Morey, Speckman, and Province (2012) and the default Bayesian *t*-tests suggested by Rouder, Speckman, Sun, Morey, and Iverson (2009). For these analyses, we report Bayes factors (*BF*) that quantify the strength of evidence in favor of the presence of an effect.[3] For example, a *BF* of 10 suggests that the data are 10 times more likely to have occurred under the model assuming the relevant effect than a model omitting this effect, whereas a *BF* of 0.10 indicates that the data are 10 times more likely to have occurred under the model omitting the relevant effect than a model including this effect.

The mean and the distribution of high probability choice proportions for each block of 50 trials and both experimental conditions are displayed in Fig. 2. A mixed model ANOVA revealed a significant main effect of trial block, $F(4.90, 235.11) = 18.42$, $p < .001$, $\eta_p^2 = .277$, $BF = 2.79 \times 10^{22}$, indicating that participants in both competitor type conditions learned across blocks.[4] These learning processes diverged adaptively depending on opponent behavior. Average choice proportions in the mimicry opponent condition accurately approached optimal probability matching ($M = .76$), whereas participants in the indifferent competitor condition approached probability maximizing ($M = .92$) by the final trial block. This main effect of competitor type was significant, $F(1, 48) = 11.66$, $p = .001$, $\eta_p^2 = .195$, $BF = 25.08$. The competitor type by trial block interaction was not significant, $F(4.90, 235.11) = 2.03$, $p = .076$, $\eta_p^2 = .041$, $BF = 0.63$.

Fig. 2 displays the full distribution of individual choice proportions alongside average response trends. Categorizing these individual choice proportions for each trial block as either probability matching (70 ± 5% of choices to the more likely outcome) or maximizing ($\geqslant$95% of choices to the more

[2] Participants also completed a short questionnaire that included the Cognitive Reflection Test (Frederick, 2005), the Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), and measures of mathematical ability. None of these measures were notably related to our dependent variables and are therefore not considered further.

[3] All Bayesian analyses were carried out in R using the BayesFactor package (v.0.9.9; Morey & Rouder, 2014). We used the anovaBF and ttestBF functions with their respective default settings with the following exception. For the anovaBF function, the number of Monte Carlo samples used to estimate *BF*s was increased to 50,000. For Bayesian ANOVAs, we computed *BF*s by contrasting the performance of a model including the relevant effect to one omitting (only) that effect. All models were constructed hierarchically such that the presence of an interaction term always involved the presence of all lower-order interactions and main effects involving the components of that higher-order interaction. For Bayesian *t*-tests, *BF*s in favor of the alternative hypothesis are reported.

[4] For all conventional ANOVAs, degrees of freedom were corrected using the Greenhouse and Geisser (1959) coefficient when the sphericity assumption had been violated.

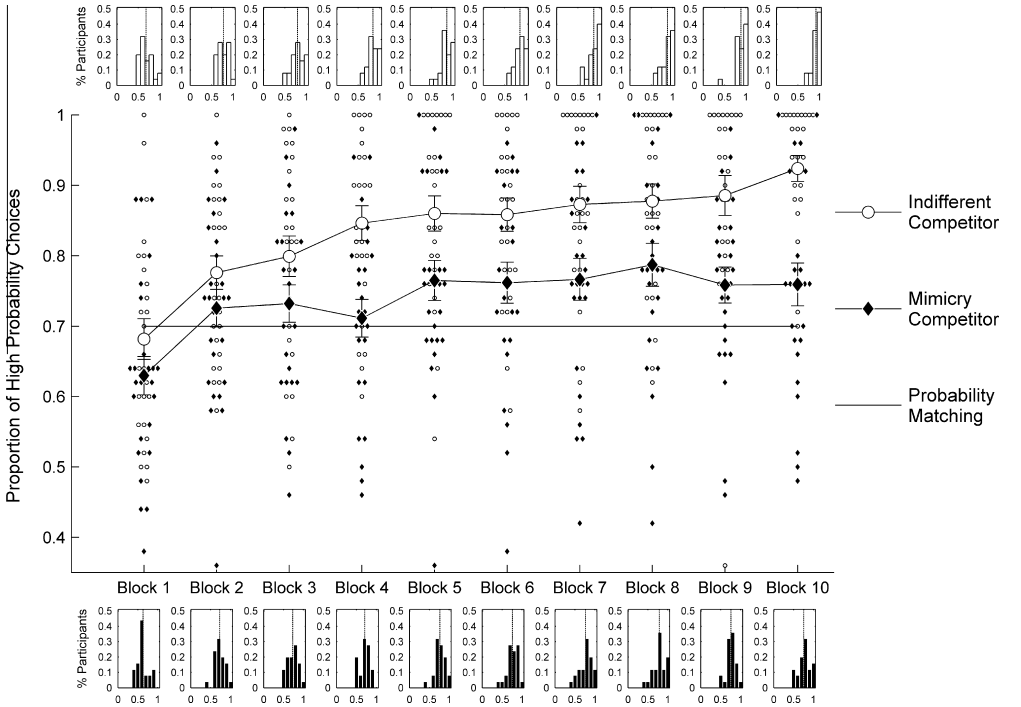

Fig. 2. Choices to the high probability option. The line graph plots the mean ± standard error proportion of choices to the more probable option averaged across participants for all 10 blocks of 50 choice trials. Probability matching is indicated by the black line at .70. Each small black diamond/white circle represents the average high probability choice proportion of one participant in the mimicry/indifferent competitor condition during a specific trial block. The bar graphs below/above the line graph show the distribution of high probability choices in the mimicry/indifferent competitor condition in percentage of participants for each trial block.

likely outcome; e.g., Newell & Rakow, 2007) revealed a similar pattern of results. Using generalized estimating equations to carry out repeated measures logistic regression analyses, we found that opponent type significantly predicted the adoption of probability matching, Wald $\chi^2(1) = 5.83$, $p = .016$, and probability maximizing, Wald $\chi^2(1) = 4.99$, $p = .025$. Participants facing imitative competitors were 2.17 times more likely to probability match; participants encountering indifferent opponents were 2.40 times more likely to probability maximize.[5]

In sum, Experiment 1 established a strong relationship between optimality and use of choice strategies in our competitive paradigm. Participants who competed against imitative opponents accurately favored probability matching; those who competed against indifferent opponents favored probability maximizing. What drove adaptive behavior in these two competitive contexts? Following the decision task, we asked participants (face-to-face) whether they "thought or were under the impression that their choices influenced the choices of the computer opponent". Less than half of participants in the indifferent competitor condition (48%) accurately disclaimed any ability to manipulate the opponent, whereas the majority of participants playing against mimicry opponents (80%)

[5] To ensure that the divergent effects of competitor type on strategy adoption ensued from the competitive nature of our task, we compared competitive probability matching and maximizing rates to non-competitive baseline rates for both strategies. These baselines were obtained from supplementary non-competitive experiments run on the same participant pool, with identical sample sizes, and analogous task features. Both competitor types provided a strategy advantage over individual (non-competitive) settings by the end of learning—in the absence of competition, participants probability maximized and matched less than participants competing against indifferent (32% non-competitive vs. 48% competitive maximizers) and imitative opponents (8% non-competitive vs. 20% competitive matchers) by the final trial block.

correctly identified their influence on opponent behavior, $\chi^2(1) = 5.47$, $p = .019$. Thus, examining people's beliefs about causal links between own and competitor behavior gave only limited indication about the motivating forces of successful strategy adoption. Therefore, we turn to an analysis of computational reinforcement learning models in an attempt to shed more light on the causes of strategy divergence.

3. Analysis of reinforcement learning models under competition

With the following computational modeling approach we aim to provide a process account for the observed differences in strategy selection under competition. Specifically, the central goal of this analysis was to illuminate which cognitive learning mechanisms participants engaged in during the task that ultimately resulted in the observed competition effect. Computational models of reinforcement learning have been shown to provide accurate and parsimonious descriptions of cognitive processes involved in similar choice settings, including learning in experimental games (Erev & Roth, 1998; Roth & Erev, 1995), learning in the Iowa Gambling task (Busemeyer & Stout, 2002; Yechiam & Busemeyer, 2005), learning in dynamic decision making tasks (Gureckis & Love, 2009), and strategy selection learning (Rieskamp & Otto, 2006). Thus, this class of models provides a suitable approach to examining learning in our paradigm and we outline potential adaptations to account for the competitive pressures introduced here.[6]

3.1. Overview and specification of learning models

We examine models of individual learning and ask whether the competitive learning requirements of our task are sufficiently represented by its competitive payoff structure. In addition, we introduce concrete conceptualizations of opponent behavior and analyze at what stage of the learning process competition may unfold its effects. All models include assumptions regarding three main components of the learning process: A utility function that specifies the goal of the learning problem, a learning rule establishing propensities for each choice option, and a choice rule that defines the course of action given current propensities (see e.g., Sutton & Barto, 1998).

3.1.1. Model 1: Individual reinforcement learning

In a learning environment where decision makers' primary goal is maximization of total payoffs, the utility of a choice is typically directly specified by its associated monetary reward (e.g., Rieskamp & Otto, 2006; Roth & Erev, 1995) as

$$u_t(i) = r_t(i). \tag{1}$$

Here, $u_t(i)$ corresponds to the utility of the monetary gains $r_t(i)$ obtained by choice i on trial t, namely, zero, two or four cents for no, split and full payoffs in our task (see Section 2.1). Based on the total reward a decision maker expects to accumulate, she establishes a propensity toward each choice, which is denoted as $q_t(i)$ for choice i on trial t. Initial propensities toward both options were assumed to equal zero, that is, $q_1(i) = q_1(j) = 0$. Following each choice, propensities are adjusted via a learning rule,

$$q_t(i) = q_{t-1}(i) + \alpha \cdot [u_t(i) - q_{t-1}(i)], \tag{2}$$

where each propensity is gradually updated in increments of the learning rate α based on the prediction error in brackets (e.g., Gureckis & Love, 2009; Yechiam & Busemeyer, 2005). Because outcomes were mutually exclusive in our task, propensities for both options are updated simultaneously

[6] We also examined models of simple heuristic decision making that presume no learning across trials. These analyses included a stochastic two-parameter version of the win-stay lose-shift heuristic—which has been shown to contribute to probability matching behavior in binary choice tasks (e.g., Gaissmaier & Schooler, 2008; Otto, Taylor, & Markman, 2011)—and a one-parameter small sample of experience model that assumes reliance on a sample of $k(t)$ past experience (see e.g., Erev & Roth, 2014). Both models provided poorer accounts of responding than a one-parameter stochastic baseline model (as indicated by negative mean differences in Akaike's Information Criterion, ΔAIC, between statistical baseline and the respective heuristic choice model; see Section 3.2) and are therefore not considered further.

regardless of the actual choice on any given trial. On the next trial, the decision maker's probability of choosing either option is given by

$$p_{t+1}(i) = \frac{e^{\theta \cdot q_t(i)}}{e^{\theta \cdot q_t(j)} + e^{\theta \cdot q_t(i)}}, \theta = 3^{10 \cdot c} - 1. \quad (3)$$

Here, the sensitivity parameter θ governs the precision with which the preferred option is chosen on the following trial (cf. Luce, 1959). If $\theta = 0$, decisions are made at random, whereas large sensitivity parameter values ($\theta \to \infty$) result in strictly deterministic choices to the option with the higher propensity. Following Yechiam and Ert (2007), an exponential transformation of θ was used to allow choice sensitivities to vary between random guessing (for $\theta \approx 0$) and fully deterministic choice (for $\theta > 700$) within narrow bounds of c, which denotes the sensitivity constant constrained between 0 and 1. Note that this individual reinforcement learning model does not include conceptualizations of the competitive nature of our task or the presence of an opposing agent beyond their associated consequences for the payoff structure.

3.1.2. Models 2–4: Competitive reinforcement learning

We extend the individualized learning model to include explicit representations of opponent behavior by adding competitive learning mechanisms to the formal model specification. This approach aims to evaluate whether adaptive decision making in our paradigm entailed responding to competitive aspects beyond the competitive payoff structure. In real-life environments, people may learn from many different competitive characteristics of a social world. Our tasks were restricted to two qualitatively different competitive pressures that reinforced divergent learning goals: Strategy mirroring opponents required consideration of ways to influence and outsmart these other agents; indifferent competitors required discounting of opposing actions. We evaluate three competitive learning models that capture these divergent influences of opponent behavior on responding with conceptually different approaches that are informed by contemporary literature on social and strategic decision making. All models rest on the assumption that the *nature* of the learning process remains unaltered by the characteristics of the competitive environment. That is to say, successful learning in each competitive context was expected to be of similar difficulty (learning rates were predicted to be equivalent) and to require similar choice precision (choice sensitivities were predicted to be comparable). Instead, the added explicit competitive mechanisms of each model were assumed to reflect the *competitive* learning mechanisms involved in our paradigm.

3.1.2.1. Model 2: Competitive payoff comparison.

In a competitive choice environment, decision makers may not perceive the utility of a choice as an absolute monetary value but as a relative concept that results from contrasting personal and rival gains. Relating to research on inequity aversion (Fehr & Schmidt, 1999), we propose a utility formation process that encompasses relative rewards based on social comparison by modifying the utility function to

$$u_t(i) = [(1 - \beta) \cdot r_t(i)] + [\beta \cdot (r_t(i) - ro_t(i))]. \quad (4)$$

Here, the utility $u_t(i)$ of choice i is expressed as the weighted sum of its absolute monetary value $r_t(i)$ (zero, two or four cents as above) and its relative benefit compared to potential opponent earnings $ro_t(i)$ (zero, two or four cents for no, split and full opponent payoffs), given the outcome and choice of the competitor on trial t. Thus, the additional free parameter β represents the weight decision makers assign to maximizing absolute payoff as compared to performing well relative to the competitor's gains. For $\beta = 0$, decision makers value their absolute payoff only, whereas for $\beta > 0$, absolute payoffs are discounted via social comparison.

3.1.2.2. Model 3: Competitive goal separation.

The focus on absolute and relative monetary choice utilities has left systematic investigations of other potential sources of choice utility largely unexplored (with few notable exceptions, e.g., Janssen & Gray, 2012; Singh, Lewis, & Barto, 2009). Yet, it is conceivable that people's choices are motivated by various other goals, for example, avoiding boredom associated with repetitive tasks (see e.g.,

Goodnow, 1955) or minimizing task completion time (Gray, Sims, Fu, & Schoelles, 2006). In our choice paradigm, two discrete learning goals are confounded when choice utilities are specified by monetary rewards only: correctly assessing the profitability of an option and attending to the competitor's choices. Incorporating this notion into the learning model, the utility function is modified to

$$u_t(i) = [(1-\beta)\cdot g_t(i)] + [\beta \cdot o_t(i)]. \quad (5)$$

Here, the utility $u_t(i)$ of a choice is expressed as the weighted sum of its accuracy $g_t(i)$ (0 for incorrect and 1 for correct guesses) and the choice of the opponent $o_t(i)$ (-1 for converging choices and 1 for diverging choices) on any given trial. The additional free parameter β represents the weight decision makers assign to choosing the correct option as compared to outsmarting the competitor in terms of choosing the opposite line of action. For $\beta = 0$, decision makers value the accuracy of their choices only, whereas for $\beta > 0$, choosing correctly and outsmarting the competitor are traded off against each other.

3.1.2.3. Model 4: Competitive strategy avoidance.

Thus far, we have considered competitive reinforcement learning models that conceptualize opponent behavior as an influence on the choice utility formation process—Model 2 and Model 3 assume that the value of a choice is directly affected by the current action of the competitor. Alternatively, individual learning about choice profitability and competitive learning about opponent strategies can be conceived as separate learning processes. This approach to learning in social contexts relates to research on social imitation learning. McElreath et al. (2005), for instance, proposed models of social learning that integrate the learning outcomes of an individual decision maker with those observed in the social environment by specifying the probability of a choice as a weighted function of its experienced success and prevalence in other decision makers. Following this notion, we propose a choice rule that separates individual and competitive motivations for each choice such that

$$p_{t+1}(i) = (1-\beta)\cdot \frac{e^{\theta \cdot q_t(i)}}{e^{\theta \cdot q_t(j)} + e^{\theta \cdot q_t(i)}} + \beta \cdot s_t(j). \quad (6)$$

Here, the probability of the next choice $p_{t+1}(i)$ is defined as the weighted sum of its relative desirability for the individual decision maker—determined by its accuracy rather than monetary profitability as discussed above—and the current preference of the opponent toward the *other* alternative $s_t(j)$, expressed as the competitor's choice proportion to option j over the past 10 trials (because feedback about prior choices remained available for a moving window of 10 choice trials; see Section 2.1). Note that, unlike in social imitation contexts, the competitive nature of our task required decision makers to avoid rather than copy opponent behavior to avoid obtaining split payoffs. Thus, the additional free parameter β represents the weight decision makers assign to maximizing choice accuracy as compared to avoiding the competitor's current choice strategy along with continuously split payoffs. For $\beta = 0$, decision makers value accurate choices only, whereas for $\beta > 0$, the desirability of a choice is discounted by its appeal to the competitor.

Table 1 summarizes the key competitive learning assumptions of all models we considered. We predicted that participants who faced indifferent competitors learned to disregard opposing actions, whereas participants who faced imitative opponents employed strategies to influence and outsmart their opponents as indicated by higher competition parameters. Thus, we predicted that $M_{\beta,mimicry} > M_{\beta,indifferent}$ and that learning models considering these competitive aspects of the paradigm would account for the behavioral data of Experiment 1 more thoroughly than the individual learning model. Moreover, the three competitive learning models will help us to disentangle at what level these competitive pressures operate: earning more money, choosing differently, or choosing strategically.

3.2. Parameter estimation and model evaluation procedure

We estimated parameters for each individual separately based on the models' accuracy in predicting the observed choice sequence one step ahead for each trial. That is, all models generate trial-by-trial choice probabilities for both response alternatives on the basis of participants' prior decisions, their associated payoffs, and the respective model's parameter values. Employing maximum

Table 1
Summary of the evaluated learning models, their complexity (as indicated by number of free model parameters), and key competitive process assumptions.

Model	*N* Parameters	Key competitive process assumptions
M1. Individual reinforcement learning	2	Competition is solely specified via the payoff structure $u_t(i) = r_t(i)$
M2. Competitive payoff comparison	3	Tradeoff between absolute and relative earnings compared to opponent's payoff $u_t(i) = [(1-\beta)\cdot r_t(i)] + [\beta\cdot(r_t(i) - ro_t(i))]$
M3. Competitive goal separation	3	Tradeoff between choosing accurately and oppositely to the competitor $u_t(i) = [(1-\beta)\cdot g_t(i)] + [\beta\cdot o_t(i)]$
M4. Competitive strategy avoidance	3	Tradeoff between choosing accurately and avoiding the competitor's strategy $u_t(i) = g_t(i),\ p_{t+1}(i) = (1-\beta)\cdot\frac{e^{\beta q_t(i)}}{e^{\beta q_t(j)}+e^{\beta q_t(i)}} + \beta\cdot s_t(j)$

Note. βs denote the added competition parameters for each competitive learning model and are assumed to capture the different strategy requirements in the mimicry and indifferent competitor conditions.

likelihood estimation we searched for the set of parameters that maximized the summed log-likelihood of the predicted choice probabilities across trials given each participant's observed responses with an iterative particle swarm optimization method (Kennedy & Eberhart, 1995).[7] The following parameter bounds constrained the optimization process: $\alpha \in [0,1]$ for the learning parameter, $c \in [0,1]$ for the transformed sensitivity θ, and $\beta \in [0,1]$ for the additional competition parameters.

The final fit of each learning model was compared to a baseline statistical model that assumes constant and statistically independent choice probabilities across trials (see e.g., Busemeyer & Stout, 2002), and hence, accounts for the data without presuming any learning. The stationary probability of choosing the more probable option pooled across all trials ($p(H), p(L) = 1 - p(H)$) is the only free parameter in this model and is estimated for each participant separately from actually observed choice proportions during the experiment. Thus, the baseline model does not automatically assume random responding or indifference between choice options but represents a valid competitor for the evaluated learning models and can provide a good account of binary choice data under the assumption that participants are not learning from obtained rewards on a trial-by-trial basis (see e.g., Gureckis & Love, 2009). To account for differences in model complexity, all learning models are evaluated by comparing differences in Akaike's Information Criterion (AIC; Akaike, 1974) between learning and baseline model for each individual, denoted as ΔAIC.[8] If a learning model is superior to the statistical baseline model and accurately describes how participants adapt their choice behavior over time, positive ΔAIC values result from this model evaluation.

3.3. Modeling results and discussion

3.3.1. Model comparison

Table 2 compares the mean ΔAIC scores for all evaluated models and indicates the percentage of participants for whom a particular learning model provides a better fit than the stationary baseline model. Highly positive average ΔAIC values demonstrate that all four learning models outperformed the stationary baseline model and provided a superior fit for the majority of participants in both experimental conditions (see Table 2). Thus, irrespective of competitor type, observed choice behavior

[7] The particle swarm optimization algorithm we used to estimate model parameters was implemented in MATLAB (Birge, 2003). For each individual, optimization proceeded iteratively with a total of 24 particles, 23 of which started at random positions while the final particle started at the best parameter combination from the previous iteration. Optimization terminated once the model fit did not improve further for at least five successive iterations. The mean (modal) number of iterations until termination was 10.7 (5), averaged across participants and all learning models considered.

[8] For extensive accounts of the use of AIC as an appropriate predictive accuracy estimation tool for the goal of model comparison see Gelman et al. (2013) and Gelman, Hwang, and Vehtari (2013).

Table 2
Mean (*SD*) differences in Akaike's Information Criterion (ΔAIC) between statistical baseline and specified model, and percentage of participants in each condition for whom the specified model provides a better fit than the baseline model (i.e., ΔAIC is positive).

	% Positive ΔAIC		Mean (*SD*) ΔAIC
	Mimicry	Indifferent	
M1. Individual RL	80%	84%	22.02 (31.30)
M2. Payoff comparison	80%	80%	26.41 (33.51) †
M3. Goal separation	80%	84%	27.15 (33.74) †
M4. Strategy avoidance	84%	84%	27.47 (32.09) †

Note. Bonferroni adjusted alpha levels of .0083 (.05/6) per test were used to determine statistical significance, i.e., †$p < .0083$.

was better accounted for when individualized or competitive reinforcement learning processes were considered.

Goodness-of-fit comparisons of all learning models amongst each other via repeated measures ANOVA revealed a significant main effect of model type on ΔAIC measures, $F(1.44, 70.71) = 8.70$, $p = .001$, $\eta_p^2 = .151$, $BF = 761.22$. Pairwise post hoc comparisons with Bonferroni corrected alpha levels (adjusted to .0083 [.05/6] per comparison) indicated that all three competitive learning models describe the observed learning processes significantly better than the individualized model: *Individual RL* vs. *Payoff Comparison,* $t(49) = -2.92$, $p = .005$, $BF = 6.50$; *Individual RL* vs. *Goal Separation,* $t(49) = -3.45$, $p = .001$, $BF = 24.66$; and *Individual RL* vs. *Strategy Avoidance,* $t(49) = -3.12$, $p = .003$, $BF = 10.69$.[9] The *Strategy Avoidance* model provides the best account of responding but the improvement relative to the other two competitive models was small and none of the competitive models fit the data significantly better than the other two (all $ps \geqslant .062$ and all $BFs \leqslant 0.82$).

3.3.2. Parameter analysis

As predicted, parameter estimates for learning rate and sensitivity constant did not differ significantly between conditions for the individual or any of the competitive learning models (all $ps \geqslant .074$ and all $BFs \leqslant 1.10$; see Table 3). This suggests that learning in both contexts was of similar difficulty and required comparable choice precision. Therefore, the individual learning model not only provided a poorer fit to the data overall, but also did not differentiate between the learning mechanisms that led to divergent choice behavior in the mimicry and indifferent opponent conditions. By contrast, the estimated additional competition parameters of all competitive learning models differed significantly between conditions (see Table 3). Each of these parameters represented the different requirements of our two environments well, while capturing different aspects of the competitive learning mechanisms participants engaged in during the task. However, because all competitive models fit the data comparably well, the relative importance of these different competitive learning mechanisms remains ambiguous. Moreover, the magnitude of learning parameter values varied considerably across models, which complicates the relative interpretability of competition parameters between models further. This is important because learning parameters may be interrelated and one estimate may influence values of other parameters. The following discussion should therefore be considered as a tentative account of possible theoretical implications of these parameter estimates for each model.

The *Payoff Comparison* model characterizes qualitatively different competitive pressures between mimicry and indifferent opponent conditions as differently weighted contrasts between personal and rival gains. Although competition parameter estimates significantly differed between experimental conditions, $t(44.50) = -2.42$, $p = .020$, $BF = 2.89$, the direction of this effect was contrary to our expectations. Participants in the indifferent condition valued comparative payoffs more strongly

[9] Follow-up analyses indicated that the improvement in fit of models with competitor consideration compared to the individual learning model did not differ between competitor type conditions; $t(48) = -0.63$, $p = .534$, $BF = 0.33$ for the *Payoff Comparison*; $t(48) = -0.19$, $p = .851$, $BF = 0.29$ for the *Goal Separation*; and $t(48) = -0.63$, $p = .529$, $BF = 0.33$ for the *Strategy Avoidance* model, respectively.

Table 3
Comparison of mean (*SD*) parameter estimates for each learning model and competitor type condition. Where estimates differed significantly between conditions, values are highlighted in bold.

	Learning (α)		Sensitivity (c)		Competition (β)	
	Mimicry	Indifferent	Mimicry	Indifferent	Mimicry	Indifferent
M1. Individual RL	.07 (.21)	.02 (.04)	.16 (.21)	.30 (.30)	–	–
M2. Payoff comparison	.08 (.22)	.07 (.14)	.13 (.18)	.20 (.26)	**.52** (.47)	**.81*** (.35)
M3. Goal separation	.08 (.21)	.07 (.14)	.42 (.38)	.28 (.22)	**.15** (.24)	**.03*** (.05)
M4. Strategy avoidance	.05 (.10)	.07 (.14)	.26 (.26)	.31 (.26)	**.08** (.11)	**.02*** (.07)

* $p < .05$.

despite their lack of influence on competitor choices.[10] The *Goal Separation* model describes the importance decision makers attribute to the choices they observe in their competitors compared to the importance they ascribe to making accurate decisions. As predicted, competition parameter estimates were significantly higher in the mimicry compared to the indifferent competitor condition, $t(26.37) = 2.44$, $p = .022$, $BF = 3.03$. This result indicates that participants in the mimicry condition forwent choices to the more probable option in favor of deviating from current opponent choices, whereas participants in the indifferent condition largely disregarded opponent choices. Finally, the *Strategy Avoidance* model assumes that the individual attractiveness of a choice is learned independently from observations of opponent behavior. Here, the competition parameter governs the weight ascribed to individual compared with competitive learning outcomes for each choice option. In line with our predictions, data from participants in the mimicry competitor condition was fit with significantly higher competition parameter estimates, $t(41.05) = 2.26$, $p = .029$, $BF = 2.16$, and therefore, these individuals showed stronger opponent strategy avoidance, than participants in the indifferent competitor group.

3.3.3. Discussion

The evaluation of reinforcement learning models suggests that the adaptive choice behavior we observed in Experiment 1 was largely driven by differences in how participants appraised opponent behavior when making decisions. The data are consistent with an account in which participants competing against imitative opponents integrated their competitors' choices in the decision process and sought to outsmart them, whereas participants facing indifferent competitors essentially disregarded their opponents' choices when deciding on a course of action. Yet, our evaluation of reinforcement learning models did not definitively explain *how* the behavior of imitative vs. indifferent competitors was integrated in participants' learning processes. The evaluated competitive learning models capture the impact of opponent behavior on responding with conceptually different approaches but account for individual choice behavior to comparable extents. Thus, further clarification of the relative merits of these competitive models is needed.

4. Experiment 2

In Experiment 1 outcome probabilities and the competitive structure of the task needed to be learned simultaneously. Therefore, the probability learning process may have hindered, confounded, or even masked the strategic learning requirements. In order to isolate learning about the competitive

[10] This seemingly puzzling finding can be explained by considering the effect of updating selected as well as unselected options. The payoff comparison process assigns negative utility to incorrect choices if the opponent selected the correct alternative and diminishes the utility of correct choices also selected by the competitor. Assuming the opponent chooses the more probable option more often than not (e.g., by probability matching or maximizing), the utility of the less probable option is learned to be negative because choices to that option likely result in incorrect predictions while the competitor accurately chooses the high probability option. This results in large relative propensities toward the high probability option and, following the ratio-of-strengths choice rule, deterministic probability maximizing (the optimal policy in the indifferent competitor condition). By contrast, if a participant disregarded comparative payoffs, the utility of the less probable option would reflect its actual expected payoff, resulting in less pronounced propensity differences between the two options. Additionally, sensitivity parameter estimates were relatively lower for this model indicating more random behavior, which may have contributed to lower estimated competition parameter values in the mimicry condition.

structure from learning about outcome contingencies, Experiment 2 replicated the basic design of Experiment 1 but used a fully described choice setting in which outcome probabilities were disclosed to participants at the start. We applied the same learning models as described for Experiment 1 and maintained identical model assumptions to allow for direct comparison between experiments.[11] Parameter estimation and model evaluation were carried out as described for Experiment 1.

4.1. Method

4.1.1. Participants

Fifty (29 female) undergraduate students from the University of New South Wales with a mean age of 19.00 years (*SD* = 2.19 years) participated in this experiment in return for course credit and performance based monetary compensation. Earnings ranged from \$2.35 to \$9.60 (1 AUD $\approx$ 0.95 USD). Twenty-five participants each were randomly assigned to the mimicry opponent (68% female) and indifferent opponent (48% female) type conditions as described for Experiment 1.

4.1.2. Dice decision task

A standard binary choice paradigm involving repeated predictions of mutually exclusive outcomes over 300 choice trials was used. Participants were asked to predict the outcome of the roll of a fair ten-sided dice with seven green and three red sides (counterbalanced across participants for red and green majority outcomes) by pressing a button labeled 'RED' or 'GREEN' displayed on a computer screen (see e.g., Newell et al., 2013). In contrast to the light bulb illumination task used in Experiment 1, the color configuration of the ten-sided die revealed outcome probabilities to participants at the start of the task. Correct choices were rewarded with four cents. Analogous to Experiment 1, each participant competed against a computerized opponent and when both decision makers converged on the correct response, the payoff was evenly split between them, that is, each agent received two cents.

4.1.3. Design and procedure

Design and procedure were identical to Experiment 1 with the following exception. The total number of choice trials was reduced to 300 because participants were not required to learn about outcome probabilities on a trial-by-trial basis, which reduced the overall complexity of the task. Thus, in Experiment 2, we used a 2×6 mixed model design with opponent type (mimicry or indifferent) as between-subjects factor and trial block (six blocks of 50 trials each) as within-subjects factor.

4.2. Results and discussion

4.2.1. Behavioral results

Fig. 3 depicts the mean and distribution of high probability choice proportions for each block of 50 trials and both experimental conditions. The figure shows that when outcome probabilities were known at the outset, participants in the indifferent opponent condition learned to probability maximize to comparable extents as under learning from experience (Experiment 1). By contrast, the adoption of optimal probability matching in the mimicry opponent condition was impaired but not eliminated.

Although outcome probabilities were known in advance, adaptive responding still required learning about the different causes of opponent behavior and we observed a significant main effect of

[11] Prior research on explicit choice advice in the Iowa Gambling task (Biele, Rieskamp, & Gonzalez, 2009) and initial strategy suggestions in fully described binary choice tasks (Newell, Koehler, James, Rakow, & van Ravenzwaaij, 2013) has incorporated the effects of advice, hints, and outcome probability information via freely estimated initial propensities to the recommended or high probability option (whereas the initial propensity for other options remained zero). Although outcome probabilities were fully disclosed in our task as well, the competitive nature of the task weakens the suitability of assumptions about variable initial propensities toward the high probability but not the alternative option. This is the case because the profitability of a choice is not directly specified by its outcome probability but contingent on expected opponent behavior. For instance, theorizing that the opponent surely will exploit the more probable option, participants might start out with high inclinations toward the less probable alternative. Therefore, we maintained all learning assumptions from the previous experiment—including the absence of initially specified choice propensities.

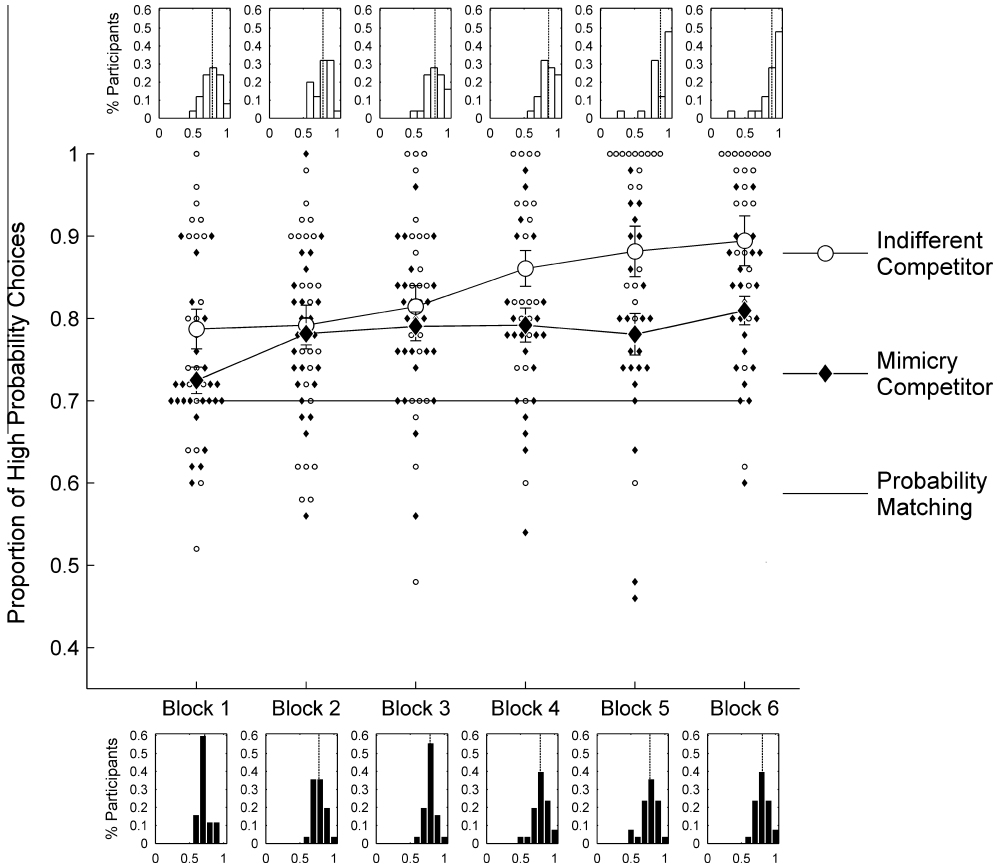

Fig. 3. Choices to the high probability option (dominant dice color). The line graph plots the mean ± standard error proportion of choices to the more probable option averaged across participants for all six blocks of 50 choice trials. Probability matching is indicated by the black line at .70. Each small black diamond/white circle represents the average high probability choice proportion of one participant in the mimicry/indifferent competitor condition during a specific trial block. The bar graphs below/above the line graph show the distribution of dominant color choices in the mimicry/indifferent competitor condition in percentage of participants for each trial block.

learning across trial blocks, $F(3.29, 157.72) = 9.78$, $p < .001$, $\eta_p^2 = .169$, $BF = 434716.13$. This trial-by-trial learning, however, led to optimal responding in the indifferent opponent condition only. Here, participants' choice behavior accurately approximated probability maximizing by the final trial block ($M = .89$). By contrast, participants in the mimicry competitor condition commenced choosing in line with optimal probability matching during the first trial block ($M = .72$) but approached over-matching by the final trial block ($M = .81$). Nevertheless, the overall competitor type effect remained and responding diverged significantly across trial blocks, $F(1, 48) = 5.16$, $p = .028$, $\eta_p^2 = .097$, $BF = 2.36$. The competition type by trial block interaction did not reach statistical significance, although the Bayesian evidence was ambiguous, $F(3.29, 157.72) = 2.49$, $p = .057$, $\eta_p^2 = .049$, $BF = 1.13$.

Analyzing individual choice proportions, again, confirmed the observed adaptive trends in aggregate responding. Using generalized estimating equations, we found that opponent type significantly predicted the adoption of probability matching, Wald $\chi^2(1) = 9.12$, $p = .003$, and probability maximizing, Wald $\chi^2(1) = 9.82$, $p = .002$. Participants playing against indifferent opponents were 6.69 times more likely to maximize; participants facing imitative competitors were 3.05 times more likely to probability match. Following the choice task, we again probed participants' beliefs about causal links

Table 4
Mean (*SD*) parameter estimates and differences in Akaike's Information Criterion (ΔAIC) between statistical baseline and specified model. Where parameter estimates differed significantly between competitor type conditions, values are highlighted in bold.

	Learning (α)		Sensitivity (c)		Competition (β)		ΔAIC	
	Mimicry	Indifferent	Mimicry	Indifferent	Mimicry	Indifferent		
M1. Individual RL	.01 (.01)	.01 (.01)	**.14** (.09)	.40** (.34)	–	–	−0.94 (17.09)	†
M2. Payoff comparison	.02 (.03)	.02 (.03)	**.11** (.09)	**.29*** (.31)	.49 (.48)	.60 (.44)	0.26 (15.26)	†
M3. Goal separation	.02 (.03)	.02 (.03)	**.28** (.26)	**.50*** (.40)	.10 (.11)	.07 (.09)	0.70 (15.25)	†
M4. Strategy avoidance	.03 (.03)	.03 (.04)	.47 (.34)	.52 (.32)	**.17** (.13)	**.10*** (.13)	3.24 (13.65)	

* $p < .05$.
** $p < .005$.
† $p < .0083$ (Bonferroni corrected alpha level).

between own and opponent choices. Although the majority of participants in the mimicry competitor condition (76%) correctly identified their influence on opponent behavior, the majority of participants interacting with indifferent competitors (60%) incorrectly endorsed being able to manipulate these opponents, $\chi^2(1) = 1.47$, $p = .225$.

4.2.2. Modeling data

Table 4 displays parameter estimates and mean ΔAIC values for all evaluated learning models. The baseline statistical model accounts reasonably well for observed choice behavior. In fact, the negative average ΔAIC value of the individual learning model (see Table 4) indicates that the baseline model outperforms this reinforcement learning based account even if differences in complexity are taken into consideration. Thus, a model of individualized learning cannot accurately describe how decision makers adapt their choice behavior over time. By contrast, complexity adjusted goodness-of-fit comparisons between the three competitive learning models and statistical baseline resulted in positive average ΔAIC scores. However, the overall performance of all evaluated learning models was lower than reported for Experiment 1. This is not surprising given that less learning was required in this 'described' version of the task.

Goodness-of-fit comparisons of all learning models amongst each other via repeated measures ANOVA revealed a significant main effect of model type on ΔAIC measures, $F(1.78, 87.15) = 13.57$, $p < .001$, $\eta_p^2 = .217$, $BF = 171681.03$. However, post hoc comparisons indicated that only the *Strategy Avoidance* model performed significantly better than the individual learning model at the Bonferroni corrected alpha level of .0083, $t(49) = -4.31$, $p < .001$, $BF = 290.50$ (see Table 4).[12] Direct comparisons of all three competitive models revealed that the *Strategy Avoidance* model significantly outperformed both the *Payoff Comparison*, $t(49) = -4.80$, $p < .001$, $BF = 1274.85$, and *Goal Separation* models, $t(49) = -3.86$, $p < .001$, $BF = 77.96$. Performance of the *Payoff Comparison* and *Goal Separation* models was not significantly different at the corrected alpha level, although the Bayesian analysis provided evidence in favor of an effect, $t(49) = -2.66$, $p = .011$, $BF = 3.60$.

As predicted, and equivalent to Experiment 1, the pace of learning remained unaffected by the competitor type manipulation and learning rates did not differ between conditions for any of the competitive models (all $ps \geqslant .425$ and all $BFs \leqslant 0.37$; see Table 4). Yet, in contrast to Experiment 1, sensitivity constants differed significantly between competitor type groups when choices were modeled by the *Payoff Comparison*, $t(28.15) = -2.82$, $p = .009$, $BF = 6.40$, or *Goal Separation* accounts, $t(41.79) = -2.22$, $p = .032$, $BF = 2.01$. Here, data from individuals in the mimicry competitor group was best fit by lower sensitivities for established choice propensities compared to the indifferent competitor condition. Thus, these models describe adoption of optimal probability matching in terms of lower choice precision. Competition parameter estimates—which describe the tradeoff between absolute and relative earnings in the *Payoff Comparison* model and the balance between choosing

[12] Follow-up analysis again indicated that the improvement in model fit of the *Strategy Avoidance* over the individual learning model did not differ between competitor type conditions, $t(48) = 1.12$, $p = .266$, $BF = 0.47$.

accurately and oppositely to the opponent in the *Goal Separation* model—did not differ between groups (all $ps \geqslant .338$ and all $BFs \leqslant 0.41$; see Table 4). By contrast, analysis of the *Strategy Avoidance* model indicates higher estimated competition parameter values for participants in the mimicry condition, suggesting that these participants assigned higher weight to avoiding the opponent's current choice strategy than participants in the indifferent opponent condition, $t(48) = 2.02$, $p = .049$, $BF = 1.46$. The estimated choice precision for this model remained unaltered by the competitor type manipulation, $t(48) = -0.54$, $p = .590$, $BF = 0.32$.

4.2.3. Discussion

Experiment 2 replicated the effect of adaptive strategy adoption under the influence of imitative and indifferent competitors within a fully described context. Overall, participants opted for high probability choices earlier on and more frequently. This finding is in line with previous research on fully described individual binary choice problems (see e.g., Newell & Rakow, 2007; Newell et al., 2013).

Removing learning requirements for outcome frequencies from the task structure diminished the overall performance of learning models compared to a stationary baseline model but led to a clearer differentiation between the competitive reinforcement learning models. The *Strategy Avoidance* model provided the best account of observed choice behavior. This model adequately reflects the different competitive requirements of our two conditions in differentially weighted efforts to avoid current opponent strategies. By comparison, the *Goal Separation* and *Payoff Comparison* models fit the data less well and explain the adoption of optimal probability matching in terms of reduced sensitivity for learned choice propensities, that is, in terms of more random choice behavior. Attributing adaptive probability matching to reduced choice rule precision, however, appears conceptually implausible because under the influence of imitative competitors participants are not less sensitive toward the profitability of choice alternatives per se. On the contrary, we suggest that it is the added requirement to outsmart the opponent that fuels optimal probability matching in this context.

5. General discussion

We evaluated individual decision making in a competitive binary choice paradigm that was based on the logic of natural foraging situations. Within this paradigm, qualitatively different competitive pressures resulted in adaptively divergent choice behavior on aggregate and individual choice levels under uncertainty (Experiment 1) and risk (Experiment 2). When competing against indifferent opponents, resources should and were found to be exploited without consideration for competitors' preferences. Much like in classic individual binary choice tasks, probability matching needed to be—and was to a large extent—dismissed as an inferior strategy. By contrast, the presence of imitative opponents required response allocations to be proportional to outcome frequencies in order to maximize payoffs. In this context, we observed an adaptive tendency toward probability matching. The application of reinforcement learning models to the data provided insight into the cognitive processes underpinning these behavioral differences.

5.1. The importance of competitive learning elements in modeling choice

Applying models of reinforcement learning highlighted two main conclusions. First, computational conceptualizations of opponent behavior are essential to explain individual choice under competition. This implication is far from trivial because the payoff structure of our paradigm already determined individual earnings competitively and may have specified competitive learning goals sufficiently. The fact that it did not, emphasizes the importance of considering a wider range of learning goals when applying models of reinforcement learning. The majority of the reinforcement learning literature, however, focusses on monetary gains for the description of learning goals and only recently have researchers begun to explore other learning goals as well (see e.g., Janssen & Gray, 2012; Singh et al., 2009). Our findings contribute to this effort.

Second, the best account of observed choices was provided by a model that conceptually separated individual learning about choice profitability from competitive learning about opponent strategies.

This separation of individual and competitive learning elements by the winning *Strategy Avoidance* model is in line with a prevalent modular view on learning mechanisms involved in both individual and social decision making. Models of strategy selection, for instance, propose a framework for individual choice based on the assumption that decision makers possess a repertoire of choice strategies that are reinforced separately in the learning process (Rieskamp & Otto, 2006). Similarly, models of social imitation learning integrate information learned by individuals with information provided by peers via separate learning processes (see e.g., McElreath et al., 2005). Our results support this approach to modeling human choice.

5.2. Implications for non-competitive binary choice

The aim of this study was to evaluate how individual binary choice is influenced by ecologically valid competitive situations. The successful adoption of optimal probability *matching* observed here, points to a potentially adaptive function of an individual choice phenomenon often dismissed as an outright violation of rational choice theory (see e.g., Koehler & James, 2009; Kogler & Kühberger, 2007; Vulkan, 2000; West & Stanovich, 2003). In this respect, our findings converge with a broader theoretical framework motivated to reveal the adaptive potential of probability matching when ecologically valid pressures are taken into account (e.g., Gaissmaier & Schooler, 2008; Gigerenzer, 2000; Green et al., 2010; Wolford et al., 2004). To accommodate our results within this framework, we would have to propose that probability matching emerges in non-competitive contexts because it is an overlearned response from frequently encountered real-world (competitive) settings. Namely, situations in which diversification of resources or spreading of choice allocations is beneficial. Previous research following this line of argument has suggested that this over- or misapplication of probability matching occurs as a reasonable response to limited and uncertain information about the true random nature of a choice task (Green et al., 2010) or because of humans' seemingly inbuilt tendency to search for patterns in sequences of outcomes (e.g., Gaissmaier & Schooler, 2008; Wolford et al., 2004). If a pattern does exist, a rule searching strategy would prove superior over static probability maximizing, and it has been shown that decision makers who probability matched in the absence of patterns were more likely to detect regularities in the outcome sequence when patterns were introduced (Gaissmaier & Schooler, 2008). This line of research emphasizes the importance of considering typical natural environmental regularities when appraising choice behavior. Our findings confirm previous suggestions (Gallistel, 1990; Gigerenzer, 2000) that the presence of competition might be another natural environmental regularity that has contributed to the persistence of probability matching in some contexts. As such, this study provides the first experimental demonstration of *optimal* human probability matching under competition in uncertain choice environments.

However, it is important not to push such a line of reasoning too far. Prior research couched firmly within the constraints of traditional (as opposed to ecological) rational choice theory has attempted to account for probability matching by identifying limitations in typically employed decision environments—e.g., lacking financial incentives or insufficient outcome feedback and training (Newell & Rakow, 2007; Shanks et al., 2002)—or by ascertaining alternative sources of utility that would motivate people to match. Goodnow (1955), for instance, noted that besides maximization of overall choice accuracy and payoffs, other sub-goals such as avoidance of boredom associated with repetitive responses or satisfaction from 'skillfully' predicting the less frequent event need to be considered to assess the utility of a choice. In other words, probability maximizing may represent a strategy that does not necessarily maximize choice utility. Moreover, the choice mechanisms underlying ecologically valid probability matching may not be as sophisticated as the successful outcome—i.e., adaptive exploitation of an environment—might suggest. Gaissmaier and Schooler (2008) and Otto et al. (2011), for instance, have shown that the same global probability matching strategy can arise from choice mechanisms that require different levels of implementation effort, including simple heuristics such as win-stay lose-shift. Although, our computational modeling analyses did not indicate a key role of simple heuristics in accounting for choice behavior (see Footnote 6), various levels of process complexity may have contributed to competitive probability matching.

The successful adoption of an optimal probability *maximizing* strategy by the majority of participants in the indifferent competitor contexts of our experiments points toward a general facilitative

effect of competitive pressures for adaptive decision making. The optimal and near optimal choice tendencies reported here were well above typically reported proportions of 'maximizers' in standard individual tasks (Vulkan, 2000) and in our own non-competitive baseline data (see Footnote 5). However, the availability of extensive outcome feedback and sufficient learning trials as provided here, have been shown to influence response optimality to comparable extents (Shanks et al., 2002). Nevertheless, the early emergence of adaptive strategy divergence within the first trial blocks (before optimality feedback was provided), confirms our claim that competition has a facilitative effect on both probability matching and maximizing.[13]

The qualitatively different competitive pressures that indifferent and imitative opponents impose on choice adequacy within our paradigm also affect the presence/absence of delayed consequences. When facing an imitative opponent, each choice is associated with delayed consequences for future reward *in addition* to immediate reward (or the absence thereof), because opponent choice is contingent on participants' prior decisions. By contrast, when faced with an indifferent opponent, each choice affects immediate reward only. In this respect, our work relates to individual choice paradigms that place immediate satisfaction in conflict with long-term goals (e.g., Herrnstein & Prelec, 1991). A key finding from this research is a kind of 'temporal myopia' where decision makers tend to ignore delayed consequences because choice is governed by local (i.e., immediate) rates of reward—a phenomenon referred to as melioration (Herrnstein & Prelec, 1991). Introducing competition to sequential choice can extend such individual temporal conflict to social settings, as is the case in our mimicry competitor condition. In contrast to frequently observed melioration behavior in individual settings, however, we found that decision makers learned to respond adequately to delayed consequences—participants probability matched adaptively when competing against imitative opponents.

5.3. Strategic binary choice under risk and uncertainty

How do these findings inform future research and relate to neighboring areas of interest? Extensions of this computerized paradigm to competitive face-to-face interactions border the intersection of research on strategic games under uncertainty and human foraging situations. Research on optimal foraging theory is primarily concerned with behavioral phenomena that emerge at group level, whereas questions regarding the adequacy of each individual's choice behavior within the group are rarely addressed. Considerations of individual strategic behavior within a social world lie at the heart of behavioral game theory. This literature predominantly concentrates on deterministic choice environments (see e.g., Camerer, 2003), whereas the role of environmental uncertainty has received less attention (but see e.g., Avrahami et al., 2005; Bereby-Meyer & Roth, 2006).

If two decision makers would interact in the competitive binary choice paradigm introduced here, their payoffs would be maximized if they diversified their choices fully between the two choice alternatives and never converged on the same option. To guarantee equal pay for both decision makers, participants would need to follow an alternating pattern whereby each decision maker exploits both options at equal rates. Thus, despite the competitive nature of this task, optimal responding would require both coordination and trust from the interacting decision makers. If, however, decision makers perceived the situation as purely competitive, the adoption of a myopic probability maximizing strategy seems tempting because, in the short run, probability maximizing avoids jeopardizing earning potential to the other player's benefit. Thus, both positive and detrimental effects of competition are conceivable in such an interactive environment and prior research on choice optimality in competitive strategic interactions has revealed both negative and facilitative influences of competition. The Dollar Auction game, for instance, provides a paradigm for irrational and paradoxical escalatory phenomena resulting from competitive dynamics (Shubik, 1971), whereas research on the perception and understanding of randomness has provided evidence for beneficial effects of competition on strategic interactions. Specifically, despite individuals' well-documented inability to generate random

[13] Additional experiments that aimed to explore different feedback and probability structure variants in competitive binary choice paradigms were run on the same participant pool and with similar sample sizes. Results from these competitive experiments indicated that the provision of outcome feedback outlining maximum rewards under an optimal strategy compared to outcome feedback alone had negligible effects on responding in similar competitive choice tasks.

series of discrete responses, fully random choice sequences can emerge in situations of competitive social interaction (Rapoport & Budescu, 1992). Thus, examination of strategic interactions under risk and uncertainty in similar sequential choice contexts would provide an interesting future avenue for this research.

5.4. Conclusion

The take-home message from the present study is that learning to choose under uncertainty and risk can be steered by competition and proceed adaptively in situations where probability maximizing *or* matching is optimal. Thus, the primary contribution of this research is that it offers a social perspective on binary choice probability learning problems that have been predominantly studied in social isolation. In taking this approach, we identified how ecologically plausible competitive pressures shape the adequacy and use of probability matching and maximizing. Moreover, we provided an experimental and computational description of optimal human probability matching under competition that contributes to our understanding of environmental determinants for adaptive human decision making.

Acknowledgments

A preliminary report of some of this work was published in the Proceedings of the 35th Annual Cognitive Science Conference (Schulze, van Ravenzwaaij, & Newell, 2013). This research was supported by Australian Research Council Grants to BRN (DP110100797 and FT110100151).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.
Avrahami, J., Güth, W., & Kareev, Y. (2005). Games of competition in a stochastic environment. *Theory and Decision, 59*(4), 255–294.
Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
Bereby-Meyer, Y., & Roth, A. E. (2006). The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. *The American Economic Review, 96*(4), 1029–1042.
Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science, 33*(2), 206–242.
Birge, B. (2003). PSOt – a particle swarm optimization toolbox for use with Matlab. In *Swarm intelligence symposium, 2003. SIS'03. Proceedings of the 2003 IEEE* (pp. 182–186).
Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253–262.
Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, N.J.: Princeton University Press.
Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*(1), 25–47.
Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review, 88*(4), 848–881.
Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences, 111*(Supplement 3), 10818–10825.
Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–128). New York: Academic Press.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.
Fretwell, S. D. (1972). *Populations in a seasonal environment*. Princeton, N.J.: Princeton University Press.
Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition, 109*(3), 416–422.
Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013a). *Bayesian data analysis* (3rd ed.). London, UK: CRC Press.
Gelman, A., Hwang, J., & Vehtari, A. (2013b). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24*(6), 997–1016.
Gigerenzer, G. (2000). *Adaptive thinking*. Oxford: Oxford University Press.
Godin, J.-G. J., & Keenleyside, M. H. A. (1984). Foraging on patchily distributed prey by a cichlid fish (Teleostei, Cichlidae): A test of the ideal free distribution theory. *Animal Behaviour, 32*(1), 120–131.
Goldstone, R. L., & Ashpole, B. C. (2004). Human foraging behavior in a virtual environment. *Psychonomic Bulletin & Review, 11*(3), 508–514.
Goodnow, J. J. (1955). Determinants of choice-distribution in two-choice situations. *The American Journal of Psychology, 68*(1), 106–116.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461–482.
Green, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proceedings of the National Academy of Sciences, 107*(37), 16401–16406.
Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95–112.
Gureckis, T. M., & Love, B. C. (2009). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. *Cognition, 113*(3), 293–313.
Harper, D. G. C. (1982). Competitive foraging in mallards: 'Ideal free' ducks. *Animal Behaviour, 30*(2), 575–584.
Herrnstein, R. J., & Prelec, D. (1991). Melioration: A theory of distributed choice. *The Journal of Economic Perspectives, 5*(3), 137–156.
Janssen, C. P., & Gray, W. D. (2012). When, what, and how much to reward in reinforcement learning-based models of cognition. *Cognitive Science, 36*(2), 333–358.
Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In IEEE, Neural Networks Council (Ed.), *IEEE international conference on neural networks proceedings* (pp. 1942–1948). Perth, Australia: IEEE.
Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition, 113*(1), 123–127.
Koehler, D. J., & James, G. (2014). Probability matching, fast and slow. In B. H. Ross (Ed.), *Psychology of learning and motivation. Psychology of learning and motivation* (Vol. 61, pp. 103–131). San Diego: Elsevier.
Kogler, C., & Kühberger, A. (2007). Dual process theories: A key for understanding the diversification bias? *Journal of Risk and Uncertainty, 34*(2), 145–154.
Kraft, J. R., & Baum, W. M. (2001). Group choice: The ideal free distribution of human social behavior. *Journal of the Experimental Analysis of Behavior, 76*(1), 21–42.
Lamb, A. E., & Ollason, J. G. (1993). Foraging wood-ants *Formica aquilonia* yarrow (hymenoptera: Formicidae) tend to adopt the ideal free distribution. *Behavioural Processes, 28*(3), 189–198.
Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.
MacArthur, R. H., & Pianka, E. R. (1966). On optimal use of a patchy environment. *The American Naturalist, 100*(916), 603–609.
Madden, G. J., Peden, B. F., & Yamaguchi, T. (2002). Human group choice: Discrete-trial and free-operant tests of the ideal free distribution. *Journal of the Experimental Analysis of Behavior, 78*(1), 1–15.
McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., Edsten, E., Efferson, C., & Paciotti, B. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior, 26*(6), 483–508.
Morey, R. D., & Rouder, J. N. (2014). *BayesFactor* (Version 0.9.9). Retrieved from <http://bayesfactorpcl.r-forge.r-project.org>.
Newell, B. R., Koehler, D. J., James, G., Rakow, T., & van Ravenzwaaij, D. (2013). Probability matching in risky choice: The interplay of feedback and strategy availability. *Memory & Cognition, 41*(3), 329–338.
Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review, 14*(6), 1133–1139.
Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition, 118*(2), 274–279.
Peterson, C. R., & Ulehla, Z. J. (1965). Sequential patterns and maximizing. *Journal of Experimental Psychology, 69*(1), 1–4.
Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121*(3), 352–363.
Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135*(2), 207–236.
Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior, 8*(1), 164–212.
Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374.
Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237.
Schulze, C., van Ravenzwaaij, D., & Newell, B. R. (2013). Match me if you can: How smart choices are fueled by competition. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 3372–3377). Austin, TX: Cognitive Science Society.
Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15*(3), 233–250.
Shubik, M. (1971). The dollar auction game: A paradox in noncooperative behavior and escalation. *Journal of Conflict Resolution, 15*(1), 109–111.
Singh, S., Lewis, R., & Barto, A. G. (2009). Where do rewards come from? In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2601–2606). Austin, TX: Cognitive Science Society.
Smith, J. M., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*(5427), 15–18.
Sokolowski, M. B. C., Tonneau, F., & Freixa i Baqué, E. (1999). The ideal free distribution in humans: An experimental test. *Psychonomic Bulletin & Review, 6*(1), 157–161.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14*(1), 101–118.
West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31*(2), 243–251.
Wolford, G., Newman, S. E., Miller, M. B., & Wig, G. S. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology, 58*(4), 221–228.
Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review, 12*(3), 387–402.
Yechiam, E., & Ert, E. (2007). Evaluating the reliance on past choices in adaptive learning models. *Journal of Mathematical Psychology, 51*(2), 75–84.