The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Melioration behaviour in the Harvard game is reduced by simplifying decision outcomes

David J. Stillwell [a]; Richard J. Tunney [a]
[a] University of Nottingham, Nottingham, UK

First Published on: 13 April 2009

To cite this Article Stillwell, David J. and Tunney, Richard J.(2009)'Melioration behaviour in the Harvard game is reduced by simplifying decision outcomes',The Quarterly Journal of Experimental Psychology,62:11,2252 — 2261

To link to this Article: DOI: 10.1080/17470210902765999

URL: http://dx.doi.org/10.1080/17470210902765999

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (11), 2252–2261

Psychology Press
Taylor & Francis Group

Melioration behaviour in the Harvard game is reduced by simplifying decision outcomes

David J. Stillwell and Richard J. Tunney
University of Nottingham, Nottingham, UK

Self-control experiments have previously been highlighted as examples of suboptimal decision making. In one such experiment, the Harvard game, participants make repeated choices between two alternatives. One alternative has a higher immediate pay-off than the other, but with repeated choices results in a lower overall pay-off. Preference for the alternative with the higher immediate pay-off seems to be impulsive and will result in a failure to maximize pay-offs. We report an experiment that modifies the Harvard game, dividing the pay-off from each choice into two separate consequences—the immediate and the historic components. Choosing the alternative with the higher immediate pay-off ends the session prematurely, leading to a loss of opportunities to earn further pay-offs and ultimately to a reduced overall pay-off. This makes it easier for participants to learn the outcomes of their actions. It also provides the opportunity for a further test of normative decision making by means of one of its most specific and paradoxical predictions—that the truly rational agent should switch from self-control to impulsivity toward the end of the experimental sessions. The finding that participants maximize their expected utility by both overcoming impulsivity and learning to switch implies that melioration behaviour is not due to the lure of impulsivity, but due to the difficulty of learning which components are included in the pay-off schedules.

Keywords: Melioration; Impulsivity; Harvard game; Decision-making; Intrapersonal externalities.

Self-control tasks have long been used as a test of rational decision making in both the human and animal literatures (Mazur, 1998; Rachlin, 2000). The Harvard game is one such task that presents the subject with a choice of (usually) two alternatives on successive trials. The pay-off from one alternative on any one trial is lower than the other alternative but in the long-term results in a higher overall pay-off. In order to maximize expected utility over the course of the task, the subject should learn to defer the initially higher reward in lieu of an even higher long-term reward. That is, a rational agent should learn to choose the option with a lower immediate pay-off.

By way of example consider an experiment reported by Herrnstein, Loewenstein, Prelec, and Vaughan (1993). They used a repeated binary choice task and constructed parallel pay-off schedules for the two options similar to that shown in Figure 1. The pay-offs for both options were

Correspondence should be addressed to Richard J. Tunney, School of Psychology, University of Nottingham, Nottingham, NG7 2RD, UK. E-mail: rjt@psychology.nottingham.ac.uk

We are grateful to Gordon Fernie for much valuable advice and discussion.

http://www.psypress.com/qjep
DOI:10.1080/17470210902765999

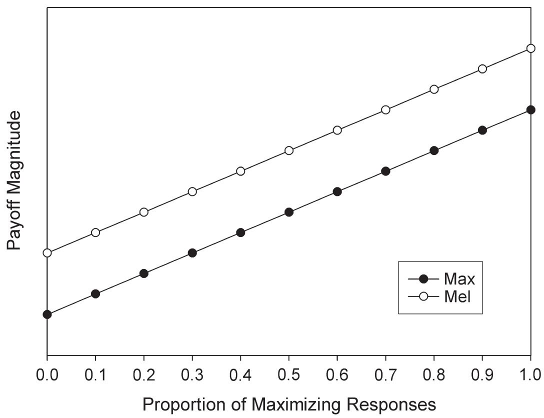

Figure 1. *Parallel pay-off schedules plotted as a function of the proportion of responses allocated to the maximizing button. The circles represent the immediate pay-off that a participant receives for choosing either the maximizing response (filled circles) or the meliorating response (unfilled circles).*

dependent upon the previous 10 choices, where choosing Option A would give a higher immediate pay-off than choosing Option B but would reduce the pay-offs of both options for the next 10 trials, and choosing Option B would give a lower immediate pay-off but would increase pay-offs over the next 10 trials. Consistently choosing Option B would ultimately maximize the overall pay-off. They found that participants tended to choose the option with the higher immediate pay-off (Option A) despite this having the effect of reducing their overall pay-off.

This and many other studies in both humans and animals have tended to paint a rather bleak picture of decision making in which impulsivity rather than self-control is the norm. Herrnstein and Vaughan (1980) suggested that the principle by which choices are made is one of *melioration*—that is, to simply select the option with the highest immediate pay-off irrespective of the consequences for future pay-offs (this is a restatement of the law of effect). Of course, in the many situations without consequences for future pay-offs this principle does maximize expected utility. However, in situations such as the one shown in Figure 1, melioration leads to choices that fail to maximize expected utility and reveals the principle to be one that violates normative expectations. Herrnstein and Prelec (1991) referred to these as intrapersonal externalities, whereby the utility of options available to the future self are not taken into account when making present decisions, in the same way that externalities in economics refer to situations in which the welfare of other individuals is not taken into account when making decisions that will affect personal welfare. Intrapersonal externalities cause a general underinvestment in activities that exhibit increasing average returns to the rate of consumption (for example, practising a musical instrument becomes more enjoyable with increased practice) and an overinvestment in activities that decrease average returns as they are consumed (such as addictive substances).

Notwithstanding the large literature documenting examples of suboptimal decision making, recent years have seen changes in our evaluation of these irrationalities. For example, the experimental economist Friedman (1998, p. 941) asserted that "every choice "anomaly" can be greatly diminished or entirely eliminated in appropriately structured learning environments" (see also Hertwig & Ortmann, 2001). The implications of this are not only that human decision making can meet the normative expectations of rational choice theory, but also that the cognitive mechanism by which the decisions are made is itself a rational process (Newell, Lagnado, & Shanks, 2007; Oaksford & Chater, 1999, 2007; Shanks, 1995), which has both the underlying capacity and the goal of maximizing expected utility, despite its cognitive constraints. It is this second implication that puts clear blue water between the evolutionary psychological view of decision making—heirs to the heuristics and biases literatures—and those sympathetic to the classical economic view of decision making.

The list of seemingly robust choice anomalies that have been seen to diminish or be eliminated in appropriately structured environments is growing. For example, preference reversals are diminished when the choices are presented to participants in the form of frequencies rather than probabilities (Tunney, 2006). Similarly, anomalies in Bayesian reasoning such as base-rate neglect and sticking in the Monty Hall dilemma are reduced when the information required to make a normative decision is presented in a frequency format (Aaron & Spivey, 1998; Gigerenzer & Hoffrage, 1995; Krauss & Wang, 2003) or is acquired through learning (Friedman, 1998; Goodie & Fantino, 1999). The phenomenon of probability matching is eliminated when the pay-offs associated with each choice consist of real, rather than facsimile, money, or participants receive meaningful feedback concerning their performance (Shanks, Tunney, & McCarthy, 2002).

With regard to melioration the tendency toward seemingly impulsive behaviour appears to be robust, despite a variety of attempts to help participants to overcome it. For instance, Herrnstein et al. (1993) provided a fairly explicit hint about how participants could maximize their pay-offs, but found that responses were only briefly improved and soon returned to suboptimal levels. Kudadjie-Gyamfi and Rachlin (1996) provided a similar hint but found no corresponding improvement at all. Warry, Remington and Sonuga-Barke (1999) attempted to reduce the motivation for participants to meliorate by reducing the immediate differential between the maximizing and meliorating options. They found that with a lower differential participants meliorated less often; however, choices were still around chance levels, and the authors noted that extrapolation of the data suggested that participants would reach asymptote at a suboptimal level. More recently, however, Tunney and Shanks (2002) were able to show that suboptimal behaviour could be overcome. They provided participants with feedback after every 100 trials detailing their score and the maximum possible score, in the same manner as Herrnstein and colleagues (1993, Experiment 3) who still found a meliorating trend, but Tunney and Shanks gave participants 1,000 trials, which proved enough for the majority of participants to learn to maximize their responses. Given this result, the irrational choices that participants often make are indicative of a failure to learn the pay-off schedules, rather than a stable decision-making bias or a failure of impulse control. In agreement with this finding, Brown and Rachlin (1999) used a task with easily imaginable pay-offs in the form of keys that opened locks. After opening each lock, the participant received a certain monetary pay-off and also a coloured key that allowed the participant to open another set of locks. Choosing the maximizing option would give a lower immediate pay-off but would leave the participant with a key that allowed a choice between more advantageous locks than those that choosing the meliorating option would have given. Brown and Rachlin found that participants tended towards the maximizing option after only 20 trials. However, since the study had just two levels of keys and locks, only the current and previous choices were taken into account when calculating each pay-off, which may not have been long enough to elicit melioration behaviour. Additionally, since all outcomes were observable, participants did not have to learn their experimentally induced preferences by experiencing the rewards of their decisions in each period, which is a central feature of most melioration experiments and seems plausible in a setting with incomplete information (Fehr & Zych, 1998).

Previously, melioration has been studied by varying either the magnitude of pay-off (e.g., Yarkoni, Braver, Gray, & Green, 2005a) or the experimental time remaining (e.g., Herrnstein et al., 1993). Participants' choices would either increase or decrease the magnitude of pay-off in current and future trials, or would alter the time remaining before the end of the experiment—in effect representing a decreasing number of opportunities to make choices and earn pay-offs. Manipulation of either of these individually tends to result in similar behaviour. The current experiment made the method by which the pay-off schedules were calculated more explicit by varying both the magnitude of the pay-off and the time remaining. The magnitude of the pay-off was dependent upon the participant's immediate choice and ignored the history of choices, whereas the time remaining depended upon the history of choices but ignored the participant's immediate choice. Therefore, by choosing the meliorating option the participant would gain a high pay-off quickly, but the experiment would ultimately end prematurely resulting in a reduced overall pay-off. In contrast, by choosing the maximizing option the participant would gain lower pay-offs but the experiment would be longer and ultimately result in a higher overall pay-off. By dividing the two components of the pay-off calculation in this way, it was expected that participants would be able to more easily learn the effects of their decisions.

An aspect of self-control tasks that has been ignored previously is that to perfectly maximize their pay-off, participants should actually switch from the maximizing option to the meliorating option at the end of the experiment. This is because the maximizing option will only maximize pay-off in the long-term, but at the end of the experiment it is better to disregard this in favour of the higher immediate meliorating reward since there is little time left to accrue later pay-offs. In real-world terms, after a lifetime of deferred gratification, upon retirement one should become both gregarious and impulsive. We examined participants' behaviour at the end of each session to test this normative expectation. If participants were shown to switch from the maximizing option to the meliorating option at the end of the experiment then it would demonstrate that they were able to successfully overcome the myopic temptation for most of the experiment and only switched when it was rationally favourable. This test essentially exploits the paradox that it is sometimes rational to be impulsive and provides a further (and we believe unique) test of normative expectations.

Method

Participants

A total of 21 undergraduate students from the University of Nottingham volunteered to take part in this experiment; 5 were male and 16 female (mean age = 21.6 years, $SD = 3.7$). Participants were not given incentives for participating in the experiment.

Pay-off schedules. Participants received points for every choice that they made, but lost game units. The experiment ended when there were no game units remaining. In the repeated binary choice task, choosing Button A over Button B returned the highest number of points per single trial; however, it rapidly reduced the number of game units remaining. Button A is therefore the meliorating option. In contrast, choosing Button B returned half the number of points per single trial but used up fewer game units, so that as long as there were more than 10 game units remaining choosing Button B would optimize participants' points pay-off. Button B is therefore the maximizing option. The magnitude of pay-off for each choice was determined by the current choice and the proportion of responses allocated to Button B over the preceding 10 trials. Figure 2 shows the pay-off schedules associated with each choice button and the effective pay-offs once the loss of game units is taken into account. Button A always gave a pay-off of 5 points, and Button B always gave a pay-off of 2.5 points. However, both buttons also reduced game units according to the formula:

Game units lost = 1 + [2 × (proportion of responses allocated to Button A in the preceding 10 trials)]

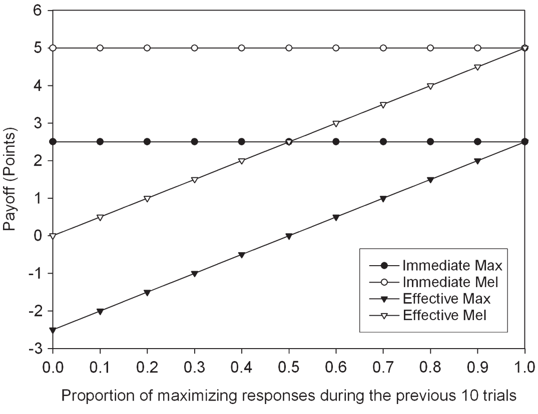

Figure 2. *Parallel pay-off schedules used in the experiment, as a function of the proportion of responses allocated to the maximizing button. The immediate pay-off schedules do not vary as a function of the proportion of maximizing responses. The effective pay-off schedules take into account the lost game units that choosing the meliorating response leads to, and they assume that each game unit lost beyond the first could have been used to gain 2.5 points by choosing the maximizing response.*

To calculate the pay-off at the beginning of each session, participants started with a history of 10 successive Button B choices. Over the 150 game units of each session, consistently choosing Button B would return a cumulative pay-off of 375 points. Consistently choosing Button A would return a cumulative pay-off of 265 points. However, the optimal solution is to switch from Button B to Button A towards the end of the session, for which a maximum pay-off of 382.5 points is possible.

The point at which it was optimal to switch from Button B to Button A at the end of each session depended upon both the number of game units remaining and the previous history of choices that a participant had made. Therefore there were actually a number of positions at which it was optimal to switch. Switching behaviour was analysed by comparing choices when there were more than 10 game units remaining to choices when there were fewer than 10 game units remaining, so the specific point at which it was optimal to switch was not crucial for our analysis of the results, since it always occurred when there were fewer than 10 game units remaining.

Stimuli

Two buttons marked "#" and "@" were displayed horizontally next to one another on the computer screen. The horizontal position (i.e., left or right) of the buttons was randomly ordered for each participant.[1] Above these two buttons, on the left side of the screen were two outcome boxes marked "Points gained on previous trial" and "Total points". On the right side of the screen were another two outcome boxes marked "Game Units lost on previous trial" and "Game Units remaining". At the top centre of the screen a horizontal bar labelled "Game Units" depicted graphically how many game units remaining there were. The colour of the bar was dependent upon the number of game units remaining; between 51 and 150 it was green, between 11 and 50 it was yellow, and between 0 and 10 it was red. Above this, another horizontal bar labelled "Points" depicted the total number of points gained during that session. This bar was based around an animated Pac-Man figure, which moved from left to right and grew larger as the total number of points increased. Participants made their choices by selecting one button or the other using the mouse.

At the end of each session, a new screen summarized the total points gained during that session and the previous sessions. The top-centre of the screen displayed textually the total points gained during the session, as well as the maximum number of points that it was possible to gain during a session. Below this, a cartoon face was presented, contingent upon whether the participant gained more points during the recently completed session than the previous session. If the participant gained more points, the face was cheerful; if an equal number of points were earned, it was neutral; and if fewer points were collected, it was dejected. Beneath these, a bar chart graphically detailed the total points gained on that session and on all previous sessions.

Design and procedure

The experiment consisted of 10 sessions, each with 150 game units. This equated to between 53 and 150 trials per session, depending upon participants' choices. At the end of each session, participants were given feedback on that session's total points compared to previous sessions and the maximum points that it was possible to gain. The experimenter remained in the testing cubicle for the first session to ensure that participants understood the task. At the end of the first session, the experimenter informed the participant that if the maximum number of points was achieved for two consecutive sessions, then the experiment would end at that time, and it would be assumed that the participant had gained the maximum number of points for all remaining sessions.

[1] Each button had a pay-off schedule associated with it (referred to as Button A and Button B in the text), as well as a label (# and @). The correspondence of each was randomly chosen by the computer for each participant.

At the beginning of the experiment participants were asked to read instructions on the computer screen (see Appendix). To initiate each trial, the points gained and game units lost outcome boxes were updated with the results of the previous trial (excluding the first trial of each session, where the outcome boxes remained blank). At the same time, two buttons were enabled marked "#" and "@", and participants were then prompted to make a choice of one of these buttons. After each choice, both buttons were disabled for between 0.5 and 1.5 seconds, and the points gained and game units lost outcome boxes were cleared to ensure that participants were aware that the outcome boxes indicated feedback from the preceding trial rather than the expected pay-off.

Since selecting Button A would lead to fewer choices, participants may have been motivated to choose this option in order to reduce the length of the experiment. To address this, the delay between trials was contingent upon the participant's game units lost on the current trial. Therefore, since all participants started with the same number of game units, the total delay over the whole experiment was similar whichever button was chosen. Between each trial, the delay in seconds was half the number of game units lost on the previous trial.

Results

The proportion of Button B responses (the long-term optimal option) was recorded across 10 sessions.

A total of 4 participants achieved the maximum score for two sessions in a row and were also able to explain how they had achieved the maximum score, and so they were excused from further sessions. For 3 of these participants, the maximum score was achieved on the eighth and ninth sessions, and for one of the participants it was achieved on the fourth and fifth sessions. In all cases, participants used exactly the same choices in the two optimal sessions, so it is reasonable to assume that had they continued until the end of the experiment they would have repeated the same choices. Therefore, in the analyses it was assumed that participants who were excused would have made the same choices in future sessions as they had made in their final session.

To analyse the data, each session was split into two blocks based on the number of game units remaining. When a participant made choices with greater than or equal to 10 game units remaining, and so should have been choosing the long-term optimal option, the choices were grouped into Block 1. When a participant made choices with fewer than 10 game units remaining, and so should be switching to the short-term optimal option, the choices were grouped into Block 2. The mean number of choices that Block 1 included were 99.3 ($SD = 30.5$) with a mean number of points gained from those choices of 305.1 ($SD = 35.1$), and the mean number of choices that Block 2 included were 5.8 ($SD = 2.1$) with a mean number of points gained from those choices of 21.1 ($SD = 5.4$).

The mean proportions of Button B responses during Block 1 of each session are shown in Figure 3 (filled circles). The Block 1 data were entered into a repeated measures analysis of variance (ANOVA) with session as the within-subjects factor.[2] This revealed a reliable effect of session, $F(4.2,\ 84.2) = 5.64$, $MSE = 0.08$, $p < .001$, $\eta_p^2 = .22$, and a reliable linear contrast indicative of an increasing trend towards maximization as the experiment progressed, $F(1, 20) = 11.56$, $MSE = 0.12$, $p < .01$, $\eta_p^2 = .37$.

If participants' behaviour is aimed at maximizing expected utility then they should never asymptote at 100% Button B responses. Instead, they ought to learn to switch to the short-term alternative toward the end of each session since this would increase their expected pay-off. The proportions of Button B choices in Block 2 of each session are shown in Figure 3 (unfilled circles) and show that

[2] The criterion for significance was set to $\alpha = .05$ for this and all subsequent tests. Degrees of freedom were adjusted using the Greenhouse–Geisser method in cases where the assumption of sphericity is violated.

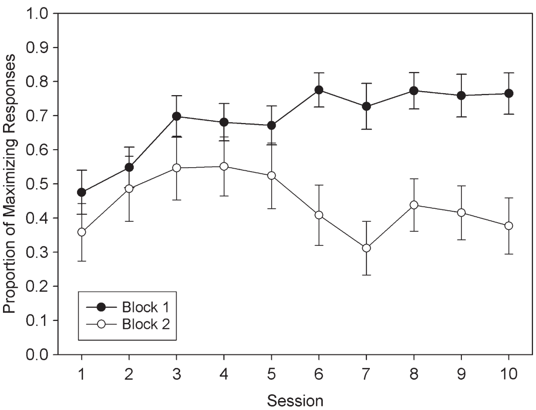

Figure 3. *Mean proportions of maximizing responses in Block 1 and Block 2 of each session. Error bars are standard errors of the mean.*

Table 1. *Comparisons showing the proportion of maximizing responses between Block 1 and Block 2 of each session, indicative of switching behaviour*

	Block				
Session	*1*	*2*	*SE*	$t(20)$	p
1	.48	.36	.06	2.07	.05
2	.55	.49	.08	0.77	.45
3	.70	.55	.08	1.79	.09
4	.68	.55	.08	1.69	.11
5	.67	.52	.09	1.72	.10
6	.78	.41	.09	4.29	<.01
7	.73	.31	.08	5.03	<.01
8	.77	.44	.09	3.65	<.01
9	.76	.42	.09	3.69	<.01
10	.77	.38	.08	4.68	<.01

toward the end of each session participants increasingly exhibit switching behaviour. To test this we compared the proportions of Button B choices in Block 1 of each session to the proportions of Button B choices in Block 2 of each session. These data were entered into a 2 × 10 repeated measures ANOVA with block and session as within-subjects factors. The ANOVA revealed an effect of block signifying that participants were indeed switching responses between Block 1 and Block 2, $F(1, 20) = 22.73$, $MSE = 0.28$, $p < .001$, $\eta_p^2 = .53$, but no effect of session, $F(5.48, 109.57) = 1.61$, $MSE = 0.17$, $p > .05$, $\eta_p^2 = .07$. The interaction between block and session was reliable, indicating that participants switched more as the experiment progressed, and they learned the relationship between choice and pay-off, $F(5.72, 114.31) = 3.92$, $MSE = 0.08$, $p < .01$, $\eta_p^2 = .16$. Pair-wise comparisons shown in Table 1 revealed that this switching behaviour became apparent during the sixth session.

Discussion

One of the fundamental assumptions of rational choice theory is that decisions maximize expected utility, which requires a cognitive decision-making system that has both the capacity and the goal of maximizing expected utility. In the Harvard game, participants have previously been observed to choose the option with a higher immediate pay-off in lieu of the option with a higher overall pay-off, leading to a suboptimal outcome (Herrnstein et al., 1993; Kudadjie-Gyamfi & Rachlin, 1996; Warry et al., 1999).

In the experiment reported here participants completed a modified version of the Harvard game in which their decisions affected the magnitude of their immediate pay-off and the number of remaining opportunities to make decisions and earn pay-offs. The results suggest that when participants make decisions they are able to take account of changes to future pay-off schedules, providing that the outcome of each decision is split into the immediate effect of the decision and the historical effects of previous decisions. By the end of the third session, which was approximately 300 trials, participants' responses were above chance. This was after fewer trials than the number found in previous self-control tasks using a 10-trial history where participants also learned to overcome their impulsivity (Tunney & Shanks, 2002).

In order to conform absolutely to normative expectations, a rational agent should ordinarily delay immediate gratification in lieu of a higher long-term reward. But if one knows that the game will end, then the rational agent should become impulsive in order to maximize their overall expected utility. One real-world analogy might be that upon retirement or diagnosis of some terminal disease, the rational agent should, after a lifetime of abstinence, embrace vice with abandon. Consistent with normative expectations, the experiment found that participants reliably switched from the choice that was optimal in the long-term to the choice that was optimal in the short-term at the end of each session, and they switched more often as the experiment progressed. This suggests that participants did not just learn that Button B was better than Button A in general, but they also learned that Button A was a better short-term option.

The principle of melioration, which states that an individual will ignore changes to future pay-off schedules that are the result of current decisions, does not account for participants' behaviour in this experiment. By dividing the outcomes of each decision so that the pay-offs were less disguised, participants took into account the consequences of their decisions over 10 sets of choice and outcome pairs. Therefore, the difficulty of optimizing behaviour in previous melioration experiments may not be due to participants not being able to take account of changes to future pay-off schedules, but instead to participants not learning to recognize at some level of understanding, either explicitly or implicitly, that this was what was necessary. In melioration experiments, the cognitive system may not be limited solely by impulsivity, but by factors such as capacity limitations or inappropriate strategy use (Yarkoni, Gray et al., 2005b).

The improved performance of participants once the pay-off calculation was split into its component parts may have implications for the treatment of behaviours with negative intrapersonal externalities such as addictions, as well as behaviours with positive intrapersonal externalities such as exercising. If one accepts that these behaviours can at least partially be modelled by temporally extended decision making, then the results suggest that one over- or underperforms these behaviours through not fully taking into account future pay-off schedule changes because they are unclear to the decision maker. Consequently, making these future effects more explicit at the time that the decision is being made might help people to incorporate them in their decision-making process and help them to maximize their welfare. The direct implication of the current experiment is that addicts should be made aware of the implication for their personal future life expectancy of each individual addictive choice. Of course, addicts are usually conscious that their addiction has negative health consequences overall; however they may find it difficult to learn the long-term impact of each individual decision on their personal welfare.

It is unlikely that the melioration model can fully explain the choices that addicts make; for example, high rates of time discounting in addicts has also been shown to play a part (Mitchell, Fields, D'Esposito, & Boettiger, 2005). Nevertheless, the results of Heyman and Dunn (2002), who found that drug-clinic patients meliorated more than control patients, suggest that addicts may be worse than others in learning to take into account the full consequences of their decisions, and so intervention at this cognitive level could be effective as part of a wider intervention that also targets other biases.

Although the points gained during the experiment were not convertible to a monetary value at the end of the experiment[3] and despite the differential between maximizing throughout a session and switching at the end of the session being small (7.5 points), the experiment seemed to provide intrinsic motivation for participants. It was set up loosely as a game, and participants anecdotally reported enjoying the challenge of attempting to solve the problem of achieving the maximum number of points possible in a session. Indeed, the finding that participants changed their behaviour during the last block of each

[3] We have since repeated this experiment using a group who earned a payment contingent upon the number of points gained in the experiment. Their performance was comparable to that of participants in the current experiment.

session in order to maximize their points would suggest that the results obtained were not simply due to a lack of motivation, as is a common criticism of decision-making research (see Hertwig & Ortmann, 2001).

It is not uncommon to demonstrate that both humans and other animals preferentially choose stimuli that are correlated with reward, for example Heyman and Tanz (1995) found that pigeons would increase their overall reinforcement rate in a version of the Harvard game by using an additional stimulus that correlated with the overall reinforcement rate. In the current experiment, the game units lost on previous trials would have been negatively correlated with the overall reinforcement rate, suggesting that perhaps participants simply used this to guide their behaviour. However, the switching behaviour observed at the end of each session suggests that participants learned the pay-off schedules more deeply than this, since they were able to appreciate one option as being optimal in the long-term and the other option as being optimal in the short-term.

To conclude, if studies of melioration are to be of use in understanding decision making or in combating addictive behaviour, what it is about melioration situations that lead participants to make suboptimal choices needs to be understood. Past literature in this area has sometimes assumed that melioration will never be overcome (Laux, 2000) or suggested that when it is overcome it is a result of impulse control (Yarkoni et al., 2005a). The results obtained in this experiment found that dividing up the pay-off schedules led to participants learning to choose according to normative expectations. This calls into question those interpretations and suggests that we should not be considering melioration in terms of an irrational decision bias that must be repressed, but instead in terms of decision making in a situation where it is difficult to learn how the pay-off schedules are calculated.

Original manuscript received 15 February 2008
Accepted revision received 25 November 2008
First published online 13 April 2009

REFERENCES

Aaron, E., & Spivey, M. (1998, August). *Frequency vs. probability formats: Framing the three doors problem.* Paper presented at the Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, University of Wisconsin, Madison WI, USA.

Brown, J., & Rachlin, H. (1999). Self-control and social cooperation. *Behavioral Processes*, *47*, 65–72.

Fehr, E., & Zych, P. K. (1998). Do addicts behave rationally? *Scandinavian Journal of Economics*, *100*, 643–662.

Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review*, *88*, 933–946.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.

Goodie, A. S., & Fantino, E. (1999). What does and does not alleviate base-rate neglect under direct experience. *Journal of Behavioral Decision Making*, *12*, 307–335.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, *6*, 149–185.

Herrnstein, R. J., & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives*, *5*, 137–156.

Herrnstein, R. J., & Vaughan, W. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior* (pp. 143–176). New York: Academic Press.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*, 383–403.

Heyman, G. M., & Dunn, B. (2002). Decision biases and persistent illicit drug use: An experimental study of distributed choice and addiction. *Drug and Alcohol Dependence*, *67*, 193–203.

Heyman, G. M., & Tanz, L. (1995). How to teach a pigeon to maximize overall reinforcement rate. *Journal of the Experimental Analysis of Behavior*, *64*, 277–297.

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, *132*, 3–22.

Kudadjie-Gyamfi, E., & Rachlin, H. (1996). Temporal patterning in choice among delayed outcomes. *Organizational Behavior and Human Decision Processes*, *65*, 61–67.

Laux, F. L. (2000). Addiction as a market failure: Using rational addiction results to justify tobacco regulation. *Journal of Health Economics*, *19*, 421–437.

Mazur, J. E. (1998). Choice and self-control. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior*. New York: Plenum Press.

Mitchell, J. M., Fields, H. L., D'Esposito, M., & Boettiger, C. A. (2005). Impulsive responding in alcoholics. *Alcoholism,Clinical and Experimental Research*, *29*, 2158–2169.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). *Straight choices: The psychology of decision making*. Hove, UK: Psychology Press.

Oaksford, M., & Chater, N. (Eds.). (1999). *Rational models of cognition*. Oxford, UK: Oxford University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Rachlin, H. (2000). *The science of self-control*. London: Harvard University Press.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology*, *48A*, 257–279.

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

Tunney, R. J. (2006). Preference reversals are diminished when gambles are presented as relative frequencies. *Quarterly Journal of Experimental Psychology*, *59*, 1516–1523.

Tunney, R. J., & Shanks, D. R. (2002). A reexamination of melioration and rational choice. *Journal of Behavioral Decision Making*, *15*, 291–311.

Warry, C. J., Remington, B., & Sonuga-Barke, E. J. S. (1999). When more means less: Factors affecting human self-control in a local versus global choice paradigm. *Learning and Motivation*, *30*, 53–73.

Yarkoni, T., Braver, T. S., Gray, J. R., & Green, L. (2005a). Prefrontal brain activity predicts temporally extended decision-making behavior. *Journal of the Experimental Analysis of Behavior*, *84*, 537–554.

Yarkoni, T., Gray, J. R., Chrastil, E. R., Barch, D. M., Green, L., & Braver, T. S. (2005b). Sustained neural activity associated with cognitive control during temporally extended decision making. *Cognitive Brain Research*, *23*, 71–84.

APPENDIX

Experimental instructions

Thank-you for agreeing to take part in this experiment. Your task is simple. You will have to repeatedly choose between two buttons, marked # and @. Simply click on a button to register your choice. As a result of your choices you will win Points. After every choice you will be shown your Points from each choice as well as your cumulative Points. As you gain more Points, Pac-Man will eat more dots and get larger! However, choices will also use up Game Units. After every choice you will be shown the Game Units used up from each choice as well as your Game Units remaining. Once these have run out then the game is over. You will play the game 10 times. Try and beat your previous score in every game! That's all there is to it—just try to win as many Points from the computer as you can before you run out of Game Units. Take as much time as you wish and please do not write anything down during the experiment.