Individuals' insight into intrapersonal externalities

David J. Stillwell[*]   Richard J. Tunney[†]

Abstract

An intrapersonal externality exists when an individual's decisions affect the outcomes of her future decisions. It can result in decreasing or increasing average returns to the rate of consumption, as occurs in addiction or exercise. Experimentation using the Harvard Game, which models intrapersonal externalities, has found differences in decision making between drug users and control subjects, leading to the argument that these externalities influence the course of illicit drug use. Nevertheless, it is unclear how participants who behave optimally conceptualise the problem. We report two experiments using a simplified Harvard Game, which tested the differences in contingency knowledge between participants who chose optimally and participants who did not. Those who demonstrated optimal performance exhibited both a pattern of correct responses and systematic errors to questions about the payoff schedules. The pattern suggested that they learned explicit knowledge of the change in reinforcement on a trail-by-trial basis. They did not have, or need, a full knowledge of the historical interaction leading to each payoff. We also found no evidence of choice differences between participants who were given a guaranteed payment and participants who were paid contingent on their performance, but those given a guaranteed payment were able to report more contingency knowledge as the experiment progressed, suggesting that they explored more rather than settling into a routine. Experiment 2 showed that using a fixed inter-trial interval did not change the results.

Keywords: intrapersonal externalities, melioration, decision-making, contingency knowledge, incentives.

1   Introduction

Experiments have long suggested that when humans make decisions they sometimes ignore even major consequences. In some situations, consequences are systematically ignored. One such situation is that of intrapersonal externalities (Herrnstein & Prelec, 1991), where changes to the utility of options available to the future self are not taken into account when making a decision, in the same way that externalities in economics refer to situations in which consequences to others are not taken into account when individuals make decisions. Overlooked intrapersonal externalities lead to an under-investment in activities that exhibit increasing average returns to the rate of consumption (for example, exercise becomes increasingly rewarding with increased practice, and also has positive effects on how rewarding other life activities are), and an over-investment in activities that exhibit decreasing average returns to the rate of consumption (use of addictive substances becomes decreasingly rewarding with increased use, and also negatively effects the perceived rewards from other life activities).

In an experiment reported by Herrnstein and colleagues (1993), participants were presented with a re-

peated binary choice task using payoff schedules similar to those shown in Figure 1. The payoff from each choice was determined by the proportion of the previous ten choices that was allocated to each option, where choosing the optimal long-term option would lead to a lower immediate payoff but would slightly increase the payoff from both options over each of the next ten trials, leading to a higher overall payoff. Conversely, choosing the optimal short-term option would lead to a higher immediate payoff but would decrease the payoff over the next ten trials, relative to the optimal option. Consistently choosing the long-term option ultimately leads to the highest payoff, although on any single trial the short-term option would give the greatest number of points. The authors found that most participants did not learn to optimise their behaviour, choosing instead the option with the greatest short-term payoff. Herrnstein and Vaughan (1980) suggested that choices are made according to the principle of melioration, in which the global rate of reinforcement is ignored.

It is not new to postulate that addicts do not fully take into account internalities, although past models have generally used hyperbolic time discounting as the theoretical reason for inconsistent preferences (Gruber & Koszegi, 2001). Nevertheless, the ecological validity of intrapersonal externalities was supported by Heyman and Dunn (2002), who found that patients recovering in drug-clinics were more likely to choose sub-optimally than control participants, suggesting that addicts may be worse than others in taking into account the full consequences

[*]Corresponding author: Psychometrics Centre, University of Cambridge, Cambridge, CB2 3RQ, United Kingdom. Email: ds617@cam.ac.uk.

[†]School of Psychology, University of Nottingham, Nottingham NG7 2RD, UK. Email: richard.tunney@nottingham.ac.uk.

Figure 1: Example payoff schedules used in the task.
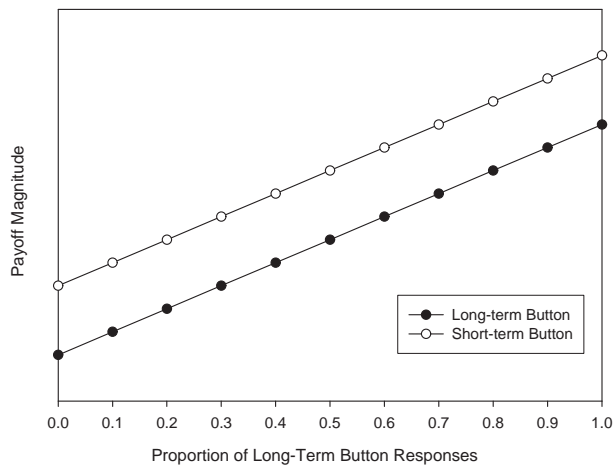
of their decisions. Further evidence for the link between intrapersonal externalities and addictions has come from neuropsychological research, which found that the level of prefrontal brain activity is associated with performance in the task (Yarkoni, Braver, Gray, & Green, 2005; Yarkoni et al., 2005). This is the same area that is implicated in studies using the Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994; Ernst et al., 2002), which is related to abuse of various substances, including alcohol and stimulants (Bechara et al., 2001). This suggests that intervention at this behavioural level could be effective in reducing addictive behaviour if a method were found to improve decisions on tasks involving intrapersonal externalities.

A series of laboratory experiments has tried without success to guide participants to choose optimally. Warry, Remington and Sonuga-Barke (1999) attempted to reduce the motivation for participants to choose suboptimally by reducing the immediate differential between the two options. They found that this helped, however by the end of their experiment participants were still choosing around chance levels and the authors noted that extrapolation of their data suggested that participants would reach asymptote at a level that was non-optimal. Two experiments have also attempted to guide participants' explicit understanding of the payoff schedules by providing a fairly explicit hint on how participants could maximize their payoffs. Herrnstein and colleagues (1993; Experiment B) found that choices were only briefly improved by the hint but soon returned to sub-optimal levels. Kudadjie-Gyamfi and Rachlin (1996) provided a similar hint, but found no corresponding improvement at all. Nevertheless, Tunney and Shanks (2002) showed that participants could overcome sub-optimal behaviour, as long as they were given regular feedback about how their behaviour compared to the optimal outcome, and

they were given around 1,000 trials to learn the schedules. This suggests that suboptimal choices in the Harvard Game are not a stable decision-making bias, but rather due to a failure to fully learn the payoff schedules.

Normally in experiments studying intrapersonal externalities participants' choices either affect the number of points gained (e.g., Yarkoni, Braver, et al., 2005) or the number of choices remaining until the end of the experiment (e.g., Herrnstein, et al., 1993). However Stillwell and Tunney (2009) modified the schedules so that both the number of points gained and the number of choices remaining until the end of the experiment were affected by participants' choices. This allowed the two outcomes from each decision to be separated so that the immediate effects were visible through the number of points gained on each trial, and the number of choices remaining until the end of the experiment decreased at differing rates depending upon the participant's history of choices. In other words, choosing myopically led to earning high payoffs through the experiment, but ultimately the experiment ended prematurely and the participant lost the opportunity to earn further payoffs. Separating the consequences from each decision made the outcomes of each decision more easily discernible, and so participants learned to choose optimally much earlier than in previous experiments. This also hints that suboptimal behaviour in the task is a failure to fully understand the payoff schedules.

In nature the outcomes from choices that are made may not be so easily divided into separate simple categories. So, if the results from laboratory intrapersonal externality experiments are to be useful in understanding the suboptimal decision-making that occurs in addictions, the process whereby participants learn to choose optimally in the simpler version needs to be understood. One process could be the result of conscious insight into the payoff schedules that participants are able to report. This mirrors research into the Iowa Gambling task which found that participants were able to explicitly report their understanding of the task (Maia & McClelland, 2004). The present experiments attempted to test explicit knowledge of the payoff schedules, to find out what participants who behave optimally are able to report about the payoff schedules. The experiments tested participants' knowledge by asking a series of questions designed to cover every scenario in the task. They used a quantitative test of participants' understanding, as these have been shown to be more sensitive than qualitative tests (Maia & McClelland, 2004).

The experiments also tested whether making participants' payment contingent on the number of points that they gained during the task had an effect on their choices. Particularly in the economics literature it is seen as crucial to incentivize participants in this way (Hertwig & Ortmann, 2001). Participants in the Contingent condi-

tion were paid based on the number of points they earned, whereas those in the Certain condition were paid a fixed amount. It is possible that giving points-contingent payments could cause participants to have more motivation and thus gain more points, or that participants would not explore as fully as they would otherwise and so would settle on a suboptimal strategy leading to fewer points (Beeler & Hunton, 1997). If, however, it is not a motivation failure that leads to poor performance on experiments using intrapersonal externalities, but rather the cognitive failure to understand the payoff schedules, then a difference might not be expected. This would suggest a cognitive component in decisions that have both long and short-term consequences that has not been fully explored.

2   Experiment 1

3   Method

3.1   Participants

Forty-nine students or staff from Nottingham University volunteered to take part in the experiment; 33 women and 16 men (mean age=27.3, SD=7.3). Participants were randomly assigned to one of two conditions; there were 26 participants in the Certain condition and 23 participants in Contingent condition. Design and procedure

Participants given standardized instructions explaining that the experiment was a decision-making experiment (see Appendix A). Also, those in the Contingent condition were told that their payment from the experiment would be based on the number of points that they gained, which would be multiplied by 0.08p/point, and those in the Certain condition were told that they would earn a guaranteed payment of £4. Pilot data had shown that the mean number of points gained over eight sessions was 5220, and so both conditions would on average earn a similar payment (minimum 4000 points = £3.20; maximum 6120 points = £4.90).

The experiment consisted of 8 sessions, each with 150 game units. This equated to between 53 and 150 choices per session, depending upon what the participant's choices were. On each trial, the points gained and game units lost outcome boxes were updated with feedback from the previous trial, and then two buttons were enabled marked '#' and '@'. Participants were then prompted to make a choice of one of these buttons. The symbol presented on each button was counterbalanced between participants, as was the payoff schedules attached to each button.

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 payoff. Choosing the short-term button used up more game units and so the experiment ended prematurely. Consequently, in order to reduce the motivation to finish the experiment quickly, choosing the short-term button increased this delay, so that the total delay over the experiment was similar whichever button was chosen. The formula used was:

$$D = 0.5 + P(short - term_{1-10}) \qquad (1)$$

Where D is the delay in seconds, and $P(short - term_{1-10})$ is the proportion of choices allocated to the short-term button over the [preceding 10 trials.

Between each session, participants were shown the points that they gained on the previous sessions. They then completed questions from four scenarios designed to probe their awareness and understanding of the payoff schedules used. The scenarios were the same each time, although they were presented in a random order after each session.

To ensure that participants understood the task, the experimenter sat with the participant for the first session, its feedback, and the first set of scenarios. Participants were then allowed to finish the other sessions in private, although the experimenter was available to answer questions.

3.2   Payoff schedules

Participants received points for every choice that they made, but lost game units. Choosing the short-term key returned 10 points per single trial, however it increased the rate at which game units were lost over the next 10 trials. In contrast, choosing the long-term key returned 5 points per single trial but used up fewer game units over the next 10 trials, so that as long as there were more than 10 game units remaining choosing the long-term key would optimise participants' points payoff. The number of game units lost after each choice was determined by the following formula:

$$GU = 1 + 2 * P(short - term_{1-10}) \qquad (2)$$

where GU is the number of game units lost, and $P(short - term_{1-10})$ is the proportion of choices allocated to the short-term button over the preceding 10 trials.

To calculate the payoff at the beginning of each session, participants started with a history of ten successive long-term button choices. Over the 150 game units of each session, consistently choosing the long-term button would return a cumulative payoff of 750 points. Consistently choosing the short-term button would return a cumulative payoff of 530 points. However, the optimal solution was to switch from the long-term to the short-term

Figure 2: Screenshot of main game screen.
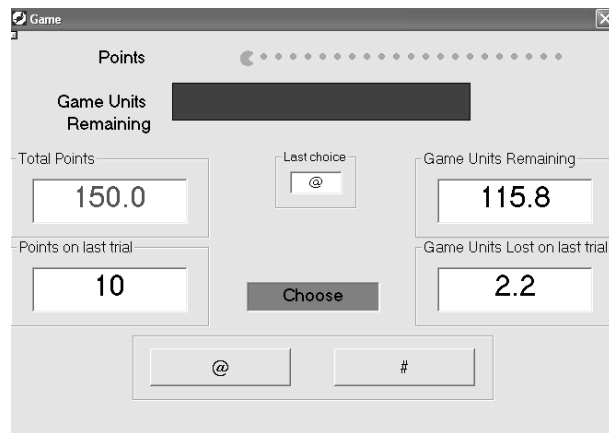
key towards the end of the session, for which a maximum payoff of 765 points was possible.

3.3 Stimuli

Figure 2 shows the main game screen. Two buttons marked '#' and '@' were displayed horizontally next to one another on the computer screen. 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 game. 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. These were designed in order to increase the saliency of the feedback. Participants made their choices by selecting one button or another using the mouse.

At the end of each session, a new screen summarised 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, and the maximum number of points that it was possible to gain during a session. Participants in the Contingent condition were also informed how much their session's points were worth monetarily. Below this, a cartoon face was presented, depending upon whether the participant gained more points during the recently completed session than the previous session. If the participant gained more

points, the face smiled; if an equal number of points were earned, it was neutral; and if a lesser number of points were collected, it frowned. Beneath these, a bar chart graphically detailed the total points gained on that session and on all previous sessions.

3.4 Contingency knowledge probe

Four scenarios were consecutively presented to each participant between each session and the next. For each scenario the participants were asked to answer how many game units would be lost and points gained if the person in it chose to continue pressing the same button (a), or what the outcomes would be if they switched to the other button (b). Participants were given a free response, and did so by typing their answer. The scenarios and their correct answers are shown in Table 1. Participants were not given any feedback on their contingency knowledge.

The points gained class of questions reflects whether participants knew that the short-term button always gave 5 points whereas the long-term button always gave 10 points. The game units lost questions reflect different levels of knowledge. Participants that answer question 1a correctly know that the lowest number of game units lost is always 1, and conversely question 4a reflects that the maximum number of game units lost is 3. Questions 1b and 2a reflect that game units lost usually increases compared to the previous trial when the short-term button is pressed, and questions 3a and 4b reflect that game units lost usually decreases compared to the previous trial when the long-term button is pressed. For these, participants do not necessarily have to understand that it is the history of choices that determines the number of game units lost, only that one button usually increases them and the other button usually decreases them. However, questions 2b and 3b are both examples of a situation where the number of game units lost does not always decrease or increase compared to the previous trial when the long-term or short-term button is pressed respectively. For a participant to correctly answer this question, it must be understood that it is the history of choices that determines the number of game units lost.

4 Results

The average number of points gained by participants in the Certain condition was 5026 (SD=451). Participants in the Contingent condition were paid based on the number of points gained during the experiment. The average number of points gained was 5021 (SD=454) leading to an average payment of £4.02 (maximum obtained: £4.75; minimum obtained: £3.45).

Table 1: Awareness probe scenarios and their correct answers. In this example the "left-hand button" refers to the long-term button, and the "right-hand button" refers to the short-term button, but the positions of these were randomized.

Scenario	Points gained	Game units lost
1. John has been choosing the left-hand button repeatedly for the last 20 turns. The last time he chose the left-hand button, he lost 1 game unit and gained 5 points.		
a) What would happen if he chose the left-hand button again?		
b) What would happen if he chose the right-hand button next time?		
1 (a)	5	1.0
1 (b)	10	1.2
2. Jane has been choosing the left-hand button repeatedly for the last 20 turns. However, 3 turns ago she switched to the right-hand button [meliorating]. The last time she chose the right-hand button, she lost 1.6 game units and gained 10 points.		
a) What would happen if she chose the right-hand button again?		
b) What would happen if she chose the left-hand button next time?		
2 (a)	10	1.8
2 (b)	5	1.6
3. Bob has been choosing the right-hand button repeatedly for the last 20 turns. However, 3 turns ago he switched to the left-hand button [maximising]. The last time he chose the left-hand button, he lost 2.4 game units and gained 5 points.		
a) What would happen if he chose the left-hand button again?		
b) What would happen if he chose the right-hand button next time?		
3 (a)	5	2.2
3 (b)	10	2.4
4. Sarah has been choosing the right-hand button repeatedly for the last 20 turns. The last time she chose the right-hand button; she lost 3 game units and gained 10 points.		
a) What would happen if she chose the right-hand button again?		
b) What would happen if she chose the left-hand button next time?		
4 (a)	10	3.0
4 (b)	5	2.8

4.1 Learning across conditions in the Simplified Harvard Game

The proportion of responses allocated to two buttons was recorded across eight sessions. Each session started with 150 game units, and was split into two blocks for purposes of analysis, based on the number of game units remaining. Any choices made while there were more than 10 game units remaining were allocated to Block 1, whereas any choices made while there were 10 or fewer game units remaining were allocated to Block 2. These two blocks represent the strategies that should be followed; for most of the game participants should choose the long-term button, but at the end of the experiment it becomes optimal to choose the short-term button. The precise optimal switching point for each session depends upon the participant's choices in Block 1, but choosing the short-term button when there were fewer than 10 game units remaining was better than choosing the long-term button. Thus, it was optimal to switch at the beginning of Block 2.

The mean proportions of responses allocated to the long-term button in Block 1 for both conditions and in each session are shown in Figure 3, and the frequencies of the proportions in each session are shown in Figure 4. The data were entered into a repeated-measures ANOVA with Session (coded numerically) as the within-subjects factor and Condition (Contingent vs. Certain) as the between-subjects factor. In this and in all further analyses, degrees of freedom were adjusted using the Greenhouse-Geisser method in cases where the assumption of sphericity is violated. The ANOVA revealed a re-

Figure 3: Proportion of responses in Block 1 allocated to the long-term button as a function of session. Error bars are standard errors of the mean.
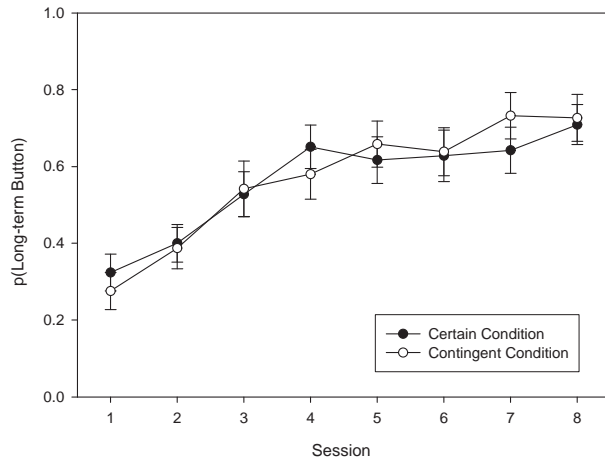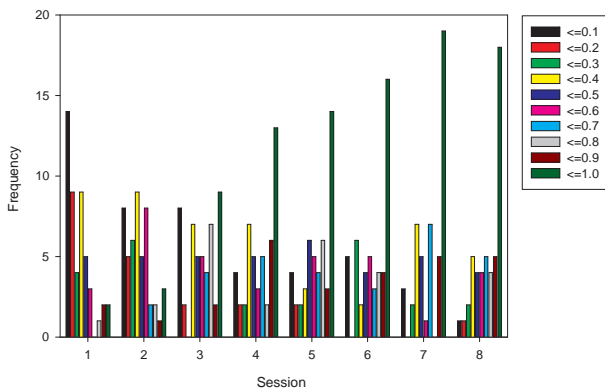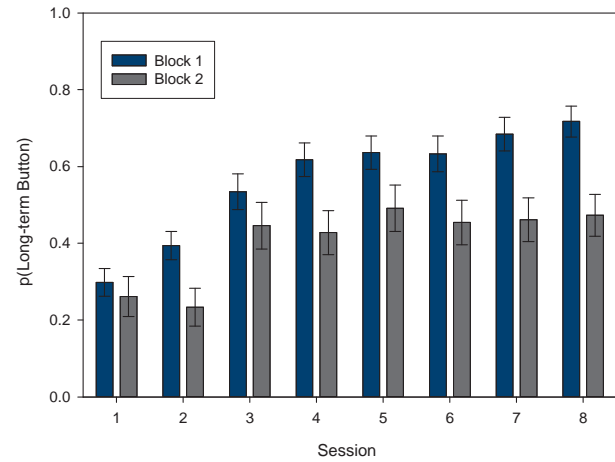

Figure 5: Proportion of long-term button choices in each session, in Block 1 compared to Block 2. Error bars are standard errors of the mean.

Figure 4: Frequencies of responses allocated to the long-term button in each of the 8 sessions of Experiment 1.

liable effect of Session ($F_{5.34, 250.9}$ = 24.31, MSE = .06, p < .001, η_p^2 = .34) and a reliable linear contrast indicative of an increasing trend towards maximization as the experiment progressed ($F_{1, 47}$ = 85.86, MSE = 0.08, p < .001, η_p^2 = .65). However, there was no evidence of an effect of Condition ($F_{1, 47}$ = .01, MSE = .40, p > .05), which suggests that rewarding participants for gaining points did not affect their performance.

If participants' behaviour is aimed at maximizing expected utility, then they should switch from the long-term to the short-term button at the beginning of Block 2. The proportions of long-term button responses for the two blocks of each session are shown in Figure 5 and show that toward the end of each session participants increasingly exhibit switching behaviour. To test this we compared the proportions of long-term button responses in Block 1 and Block 2 of each session. These data were entered into a 2x8x2 repeated-measures ANOVA with Session and Block as within-subjects factors, and Condition

as the between-subjects factor. The ANOVA showed an effect of Block, signifying that participants were switching responses at the end of each session ($F_{1, 47}$ = 30.17, MSE = .16, p < 0.001, η_p^2 = .39). A reliable interaction was found between Session and Block ($F_{5.4, 252.3}$ = 3.05, MSE = .04, p < 0.01, η_p^2 = .06). The linear trend across Sessions differed between Blocks ($F_{1,47}$ = 8.91, MSE = .06, P < .01, η_p^2 = .16) revealing that participants' choices between Block 1 and Block 2 increasingly diverged as the experiment progressed, and pair-wise comparisons revealed that across the 8 sessions participants increasingly switched from the long-term button in Block 1 to the short-term button in Block 2. This switching behaviour became consistently apparent after the fourth session. There was no reliable between-subjects main effect of Condition, nor any interactions between Condition and Block or Session, suggesting that rewarding participants for gaining points did not mediate their switching behaviour.

4.2 Probe responses indicative of optimal behaviour

Between each session and the next, participants answered questions from four scenarios designed to test their knowledge of the payoff schedules used in the experiment. One type of question asked how many points would be gained after the next choice was made. However, ceiling effects were found as most people answered the questions correctly even in the early sessions. Therefore, as this type of question was not able to discriminate between individuals, they were not analysed further.

The second type of question asked how many game units would be lost after the next choice was made. The

Table 2: Percentage of times that answer is given in Experiment 1, and whether answer is associated with long-term choices in Block 1 of the next Session.

	Q1A Corr.	Q1B Corr.	Q2A Corr.	Q2B Corr.	Q3A Corr.	Q3B Corr.	Q4A Corr.	Q4B Corr.	Q3B Err.[b]	Q4A Err.[b]
% chosen	73%	51%	57%	40%	39%	30%	59%	38%	35%	28%[a]
Δ % chosen: session 8 minus session 2[a]	37%**	41%**	39%**	27%**	10%	4%	4%	10%	24%*	4%
Fixed effect (t)	3.56**	3.79**	2.73**	-0.12	-0.14	-2.11*	-1.3	0.57	3.72**	0.1

Evidence of learning across all sessions was tested using a within-subjects ANOVA with the 7 payoff contingency probes as the within subjects factor and the proportion of correct responses as the dependent variable.
[b] The incorrect answer for Q3B was 2.6, and for Q4A it was 3.2. Both of these errors are participants incorrectly assuming that the game units lost would increase from the previous trial.
* = p<.05; ** = p<.01

percentage of participants who answered each question correctly is presented in Table 2, as well as an analysis of whether participants learned to answer the question correctly as the sessions progressed. In order to test how effective these were in relating the knowledge that answering the question correctly represented to behaviour in the experiment, participants' correct or incorrect answers to the probe questions were used to predict performance in the next session. Mixed effects models using the nested data, with response to the probe question (correct or incorrect) as a first level predictor and participant ID as the grouping variable, found significant positive effects for Q1A, Q1B and Q2A. To our surprise, it also found a significant negative correlation for Q3B and a sizable (but non-significant) correlation for Q4A, and so these were explored further. It was found that participants who learned to choose the long-term button during Block 1 made systematic errors on Q3B and Q4A; incorrectly assuming that the game units lost would increase from the previous trial (in Q3B, from 2.4 to 2.6, and in Q4A from 3 to 3.2) when it would not. In fact, further mixed effects models found that the misconception in Q3B was associated with optimal behaviour in Block 1.

The systematically incorrect answers by participants who performed optimally in Block 1 of their next game go some way to explaining the pattern of correlations. In order to answer questions 1B and 2A correctly, which participants who chose optimally were more likely to do, it is necessary to know that usually pressing the short-term key leads to an increase in the number of game units lost compared to the previous trial. It is perhaps not surprising that there is such a close relationship between answering question 1A correctly and choosing optimally. Participants who chose optimally would have extensive experience of scenario one's position on of the payoff schedules. It is noteworthy that 3A and 4B were not associated with optimal decision making. These reflect a

mistaken understanding that pressing the short-term button always leads to an increase in the number of game units lost on the next trial. Due to the 10 trial history used when calculating the payoff schedules this was not always the case, but as a heuristic it is correct more often than not.

The overall understanding of the payoff schedules, using the proportion of correct answers across all eight questions as the dependent variable, was compared between the Certain condition and Contingent condition using a 7x8x2 mixed ANOVA with Session and Question as within-subjects factors, and Condition as a between groups factor. The analysis did not find a reliable main effect of Condition (F1, 47 = 2.05, MSE = 3.99, p = .16), nor did it find an interaction between Condition and Question (F4.5, 213 = .47, MSE = .45, p = .78) suggesting that both groups performed equally well on each question, however it did find a reliable interaction between Session and Condition, (F3.54, 166.5 = 2.76, MSE = .78, p = .04, η_p^2 = .06). The linear contrast for the Session and Condition interaction (F1,47 = 5.20, MSE = .36, P = .03, η_p^2 = .10) suggests that despite the lack of a main effect, participants in the Certain condition learned the correct answers to questions at a faster rate compared to participants in the Contingent condition. It can be seen from Figure 6 that participants who were paid contingent on their performance stopped improving their understanding of the payoff schedules after the second game, whereas those who were given a guaranteed payment continued to improve beyond this.

In conclusion, participants in both conditions learned to optimize their behaviour and even to switch towards the immediately beneficial option towards the end of the experiment. There was however no evidence that paying participants contingently on their choices changed their choice behaviour, although despite this, those who were given a certain payment for participation benefitted from

Figure 6: Proportion of correct answers to contingency probe questions in each session, separated by condition. Error bars are standard errors of the mean.
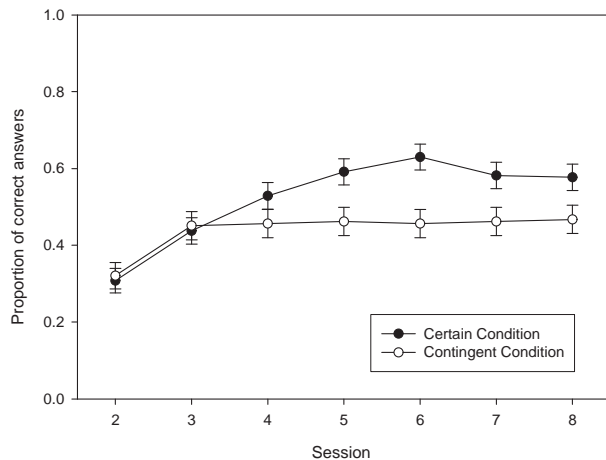
extended learning of the payoff schedules and were ultimately able to better predict the outcome of choices on the task. There is some indication that participants who performed optimally did not form a full understanding of the historical interaction between the two options affecting the number of game units lost. Instead they generalised that the short-term key increased the number of game units lost compared to the previous trial, leading to systematically incorrect answers on some questions. This may explain why the Contingent condition had a poorer overall understanding of the task but still performed equally well as the Contingent condition.

5   Experiment 2

Experiment 1 modified the Simplified Harvard Game by providing knowledge probe questions. It is possible that by asking these questions they changed participants' learning about the task payoffs. In order to test this, the experiment was repeated with two new groups. One group received the knowledge probes throughout, and the other group received knowledge probes only for the final three sessions. The previous experiment confounds the wait between trials and performance on the task, such that acting impulsively leads to a longer wait between trials. This had the advantage that the overall length of time on the task would be equal for both groups, as those who behave suboptimally end the experiment after fewer trials but have to wait longer between trials to make up for it. Unfortunately this also means that participants could use the waiting time between trials to gauge how well they were doing on the task. Therefore, in Experiment 2 the inter-trial interval was set to a fixed time rather than based on the history of choices.

6   Method

6.1   Participants

Forty-eight students from Nottingham University volunteered to take part in the experiment; 33 women and 15 men (mean age=25.0, SD=6.3). Participants were assigned to one of four conditions based on the order in which they presented.

6.2   Design and procedure

The design was largely identical to the first experiment, with half of the participants paid contingent on their performance and half paid a guaranteed amount. However, an additional independent variable was added, orthogonal to the first, which was the point at which contingency knowledge probe questions were asked; half the participants were asked after each game, and the other half were asked only after the final three games. Additionally, the inter-trial interval was set to 1 second, rather than the variable interval as in the previous experiment.

6.3   Contingency Knowledge Probe

In order to simplify the probe questions, participants' answers were restricted to whether the number of game units lost "would stay the same", "would increase" or "would decrease". Otherwise, the questions and were identical to the game unit questions from the previous experiment, shown in Table 1.

7   Results

The average number of points gained by participants in the Certain condition was 5297 (SD=306). Participants in the Contingent condition were paid based on the number of points gained during the experiment. The average number of points gained was 5383 (SD=342) leading to an average payment of £4.31 (maximum obtained: £4.72; minimum obtained: £3.67).

7.1   Learning across conditions in the Simplified Harvard Game

The mean proportions of responses allocated to the long-term button in Block 1 for the four conditions and in each session are shown in Figure 7, and the frequencies of the proportions in each session are shown in Figure 8. The data were entered into an 8x2x2 repeated-measures ANOVA with Session (1 to 8) as the within-subjects factor (coded numerically) and the two Conditions (contingent vs. certain payment, and early vs. late knowledge probe) as between-subjects factors. The ANOVA re-
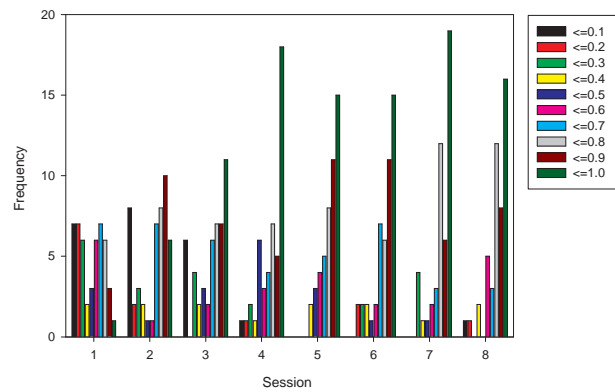
Figure 7: Proportion of responses in Block 1 allocated to the long-term button as a function of session. Error bars are standard errors of the mean.

Figure 8: Frequencies of responses allocated to the long-term button in each of the 8 sessions of Experiment 2.
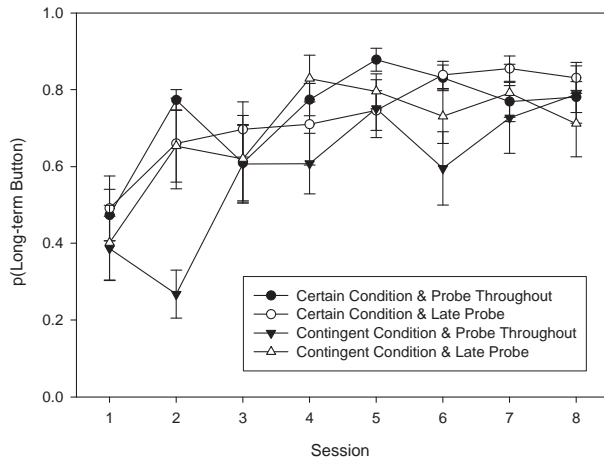
vealed a reliable effect of Session ($F_{3.49, 153.42}$ = 16.08, MSE = .09, p < .001, η_p^2 = .27) and a reliable linear contrast indicative of an increasing trend towards maximization as the experiment progressed ($F_{1, 44}$ = 42.9, MSE = .10, p < .001, η_p^2 = .49). The effect of Payment Condition ($F_{1, 44}$ = 1.01, MSE = .21, p > .05) was not significant, suggesting that rewarding participants for gaining points did not affect their task performance. For Probe Condition ($F_{1, 44}$ = 3.85, MSE = .21) the effect was almost significant. To properly analyse this, we compared the two probe conditions for Sessions 2-6, as neither group had received the probe during the first session, and both groups had received the probe during Sessions 7 and 8. The 5x2 within-subjects ANOVA found an effect of Probe Condition ($F_{1, 46}$ = 4.97, MSE = .14, p = .031) indicating that the group who received the probes chose the long-term button more often than the group who did not. One explanation for this finding is that asking knowledge questions encouraged participants to gain explicit knowledge of the payoff schedules, which in turn improved their performance on the task.

If participants' behaviour is aimed at maximizing expected utility, then they should switch from the long-term to the short-term button during the final block. To test this we compared the proportions of long-term button responses in Block 1 and Block 2 of each session. These data were entered into an 8x2x2x2 repeated-measures ANOVA with Session (1 to 8) and Block (1 to 2) as within-subjects factors, and the two Conditions (contingent vs. certain payment, and early vs. late knowledge probe) as the between-subjects factors. The ANOVA showed an effect of Block, signifying that participants were switching responses at the end of each session ($F_{1, 44}$ = 58.34, MSE = .21, p < 0.001, η_p^2 = .57). A reli-

able interaction was found between Session and Block, signifying that participants switched more as the experiment progressed ($F_{4.4, 192}$ = 7.90, MSE = .05, p < 0.01, η_p^2 = .15). The linear contrast for the Session and Block interaction ($F_{1, 44}$ = 29.74, MSE = .04, P < .001, η_p^2 = .40) revealed that participants switched more as the experiment progressed, and pair-wise comparisons revealed that this switching behaviour became consistent from the second session onwards. There was no reliable interaction between the two Conditions and Block or Session, suggesting that rewarding participants for gaining points did not mediate their switching behaviour.

7.2 Probe responses indicative of optimal behaviour

The percentage of participants who answered each probe question correctly is presented in Table 3. As in Experiment 1, mixed effects models were used to test for a relationship between responses to the probe questions and proportion of long-term choices in Block 1 of the next Session, while holding the participant ID constant in order to account for the different skill levels amongst participants. A significant positive relationship was found between both Q1A and Q2A with Block 1 of the next session; participants who answered those questions correctly were more likely to choose the long-term button more often in Block 1 of the next Session. As in Experiment 1, correct answers for Q3B and Q4A were negatively related to performance in the next ask (although not significantly in this experiment). In order to be consistent with the previous experiment, the equivalent incorrect answers for these questions were analysed. The direction of the effect was the same as in Experiment 1; participants who behaved optimally incorrectly assumed that the rate that game units were lost would increase from the previous trial, although this was not a significant effect.

Table 3: Percentage of times that answer is given in Experiment 1, and whether answer is associated with long-term choices in Block 1 of the next Session

	Q1A *Corr.*	*Q1B* *Corr.*	*Q2A* *Corr.*	*Q2B* *Corr.*	*Q3A* *Corr.*	*Q3B* *Corr.*	*Q4A* *Corr.*	*Q4B* *Corr.*	*Q3B* *Err.*[a]	*Q4A* *Err.*[a]
% Chosen	74%	58%	68%	46%	39%	34%	57%	46%	40%	33%
Fixed effect (t)	3.02**	2.60*	1.44	0.32	0.77	-0.78	-0.2	0.86	1.08	0.52

[a] The incorrect answer for both Q3B and Q4A was that the game units "would increase".

* = p<.05; ** = p<.01

The smaller number of significant effects in this experiment may be due to the format of the probe questions. In the previous experiment, participants were given a free response, so guesses were unlikely to be correct by chance. But in the current experiment, participants were given a multiple response option, so guesses were more likely to be correct.

The overall understanding of the payoff schedules, using the proportion of correct answers across all eight questions as the dependent variable, was compared between the Certain condition and the Contingent condition using a mean of the final three sessions where all participants answered contingency knowledge probe questions. A 3x8x2x2 mixed ANOVA with Session (6-8) and Question (q1a to q4b) as the within-subjects factors, and Probe Condition and Payment Condition as between groups factors. The analysis found a reliable main effect of Payment Condition (F1, 44 = 4.40, MSE = 1.17, p = .04, η_p^2 =.09), indicating that those who were given a certain payment answered more questions correctly in the final three sessions than those given a contingent payment. There was no main effect of Probe Condition (F1, 44 = .06, MSE = 1.17, p = .81) indicating that the timing of the probes did not affect participants' understanding of the task. It can be seen from Figure 9 that participants who were paid contingent on their performance generally stopped improving their understanding of the payoff schedules after the second session, whereas those who were given a certain payment continued to improve beyond this, although they took longer before they reached their optimal understanding of the task.

In conclusion, Experiment 2 replicated the behavioural and contingency knowledge results of Experiment 1 and there was no evidence that the probe questions changed participants' behaviour in the task or their understanding. Participants optimized their behaviour whether they were paid contingent upon their performance or not, and those who performed the best seemed to follow a heuristic that did not fully characterise the complexity of the payoff schedules. There was also evidence that those paid contingently on their responses did not learn as much explicit

Figure 9: Proportion of correct answers to contingency probe questions in each session, separated by condition. Error bars are standard errors of the mean.
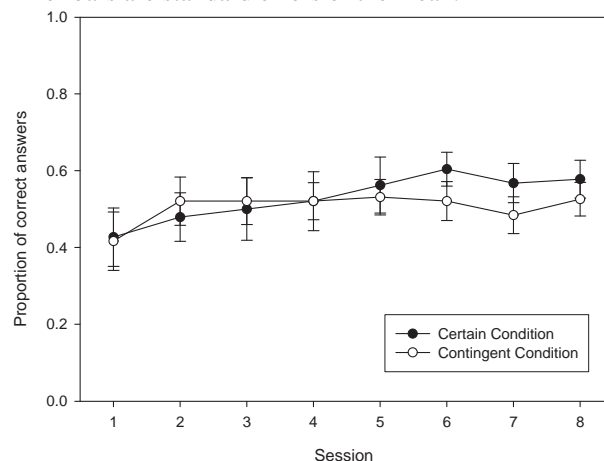
knowledge about the payoff schedules. Finally, there was no evidence that fixing the inter-trial interval altered participants' performance in the task.

8 Discussion

In the Harvard Game experiments reported here, participants learned to take account of the intrapersonal externalities inherent in the task, maximizing their expected utility. By the final session, most participants chose the long-term option for the majority of the session, and switched responses to the short-term option towards the end. In real-world terms, participants learned to choose activities that would increase their long-term welfare rather than those that gave immediate gratification, but once they realised that the end was close participants learned to prioritise their short-term needs.

By asking participants to complete quantitative questions about their predictions in the experiment we could distinguish between the different conceptualisations of the task that participants held. Based on the questions that

participants who made more optimal choices answered correctly, and the errors that participants made, we found evidence that participants appeared to use a generalised heuristic that one option would usually increase the rate at which choice opportunities remaining in the experiment decreased. There was no evidence that participants who made more optimal choices realised that the other option would usually decrease the rate at which the number of choice opportunities remaining in the experiment decreased.

In addition, the experiments found that paying participants based on their choices had no observable effect on their choice behaviour, despite the claims of Hertwig and Ortmann (2001). The finding that payment type does not affect choice behaviour is consistent with similar research using the Iowa Gambling Task (Bowman & Turnbull, 2003) which found no difference between real or facsimile money.

It is possible that contingent payments increased participants' motivation but this was offset by a decrease in exploratory behaviour. Supporting evidence for this was found from the analysis of how participants' understanding of the payoff schedules changed across the eight sessions. In the group who were paid contingent on their performance, their ability to successfully predict the outcome of choices in the simplified Harvard Game plateaued after the second game, whereas the group who were given a guaranteed payment continued to improve beyond this. In both experiments, the participants who were paid a fixed amount learned more about the payoff schedules by the end of the experiment than participants who were paid contingent upon their performance. This is consistent with previous research by Schwartz (1982), who found that participants' learning of complex sequences was impaired by giving them contingent reinforcement. Schwartz (1982) concluded that participants repeated what worked in the past rather than trying to understand the task. Since explicit knowledge of the payoff schedules is overall related to performance on the task, it is an open question whether, given enough trials, participants not paid contingent on their reinforcement would ultimately understand the task better and so also learn to perform better than those paid contingent on their performance, or whether participants paid contingent on their performance would extend their initial exploratory period so that they gained more understanding before settling into a pattern of responses that seems to work.

The experiments provide insight into what people who take into account intrapersonal externalities understand. As we found evidence that explicit understanding is related to performance in the task, but found no difference in performance based on how participants were paid, this suggests that the task is predominantly a cognitive one. It is possible that an intervention could be piloted to increase the occurrence of behaviours with positive intrapersonal externalities, or decrease the occurrence of behaviours with negative intrapersonal externalities. An intervention could use pervasive digital devices to provide immediate and personalized feedback each time an individual engages in behaviour with intrapersonal externalities.

As well as intrapersonal externalities, other factors are associated with apparently impulsive behaviour. For example high rates of time-discounting are also related to addictive behaviour (Mitchell, Fields, D'Esposito, & Boettiger, 2005; Kirby, Petry & Bickel, 1999; Vuchinich & Simpson, 1998). The difficulty that the human decision-making system has in taking account of intrapersonal externalities should be considered as an additional factor leading to addictive behaviour.

To conclude, if the results of experiments on intrapersonal externalities are to be useful in understanding suboptimal behaviour in the real world, the differences in understanding between participants who learn to behave optimally and those who do not needs to be understood. Our results suggest that explicit awareness is useful for making optimal decisions in the simplified Harvard Game, but that participants who make more optimal choices do not have, and do not need, a full knowledge of the historical interaction that leads to each payoff. Instead participants learn a simpler conception which emphasises that one option generally seems to make an aspect of their situation worse than it was previously, on a myopic choice by choice basis. When the two aspects (history of choices and current choice) are combined into a single outcome in the full Harvard Game, and presumably in real life intrapersonal decision-making, it is much more difficult to learn this relationship and this may explain why the simplified Harvard Game is simpler to learn.

References

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

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S. W., & Nathan, P. E. (2001). Decision-malting deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia, 39,* 376-389.

Beeler, J. D., & Hunton, J. E. (1997). The influence of compensation method and disclosure level on information search strategy and escalation of commitment. *Journal of Behavioral Decision Making, 10,* 77-91.

Bowman, C. H., & Turnbull, O. H. (2003). Real versus facsimile reinforcers on the Iowa Gambling Task.

Brain and Cognition, 53, 207-210.

Ernst, M., Bolla, K., Mouratidis, M., Contoreggi, C., Matochik, J. A., Kurian, V., et al. (2002). Decision-making in a risk-taking task: A PET study. *Neuropsychopharmacology, 26,* 682-691.

Gruber, J., & Koszegi, B. (2001). Is addiction "rational"? Theory and evidence. *Quarterly Journal of Economics, 116,* 1261-1303.

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.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128,* 78-87.

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

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

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.

Schwartz, B. (1982) Reinforcement-induced behavioural stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General 111,* 23-59.

Stillwell, D., & Tunney, R. J. (2009) Melioration behaviour in the Harvard Game is reduced by simplifying decision outcomes. *Quarterly Journal of Experimental Psychology 62,* 2252-2261.

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

Vuchinich, R. E., & Simpson, C. A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology, 6,* 292-305.

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. (2005). 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. (2005). Sustained neural activity associated with cognitive control during temporally extended decision making. *Cognitive Brain Research, 23,* 71-84.

Appendix: Instructions to participants

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 with the mouse 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, Pacman 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 8 times. Try and beat your previous score in every game!

At the end of every game, you will be asked a series of questions relating to potential scenarios within the game. Your answers to these questions will have no effect on the games that you play.

Your payment from this experiment will be £4.

[Participants in condition contingent were shown the following text in replace of the previous paragraph]

Your payment from this experiment will be based on the number of points that you gain during the games. This will be calculated on the basis of 0.08p/point. After each game, you will be shown your current earnings so far.

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.