



AUC Undergraduate Journal of Liberal Arts & Sciences

Capstone Issue Vol. 1 2012

AUC Undergraduate Journal of Liberal Arts & Sciences Capstone Issue Vol. 1 2012

www.auc.nl

**AUC Undergraduate Journal
of Liberal Arts & Sciences
Capstone Issue Vol. 1 2012**

Copy Editing

Zuzanna Fiminska
Simon Hadlich
Dr. Rebecca Lindner
lanthe Schepel

Images

Rocky Lubbers, Money Crush. Uploaded September 29, 2011 via Flickr. Creative Commons Attribution.
'Truus, Bob & Jan Too!', Back to Back Vigelandsanlegget. Uploaded October 20, 2010 via Flickr. Creative Commons Attribution. Permission granted by Vigeland Museum, Norway.
Courtney Benefiel, Ride to Remedy. Uploaded February 6, 2008 at <http://ridetoremedy.com>.

The AUC Undergraduate Journal of Liberal Arts and Sciences is a biannual, interdisciplinary publication showcasing outstanding undergraduate academic papers. The Journal aims to demonstrate the strength of undergraduate scholarship at AUC, to reflect the intellectual diversity of its academic programme, to encourage best research and writing practices, to facilitate collaboration between students and faculty across the curriculum, and to provide students with opportunities to gain experience in academic reviewing, editing and publishing.

FOREWORD

This inaugural issue of the AUC Undergraduate Journal of Liberal Arts and Sciences takes its inspiration from the Capstone projects that all of our students complete in the final year of their degree. The Capstone, an extended research thesis that requires significant independent study on a substantial scholarly topic within a student's major, has yielded some of the strongest academic work at AUC.

Many students produced remarkable, often interdisciplinary projects, addressing exciting and pertinent research questions facing contemporary science, culture, and society. This reflects the aims of AUC to foster an environment that encourages and enables high quality undergraduate scholarship that works through and across disciplinary boundaries. The strength of our students' Capstone work is also encouraging because, looking ahead to the coming years, it suggests that this journal will have more than enough outstanding material to draw upon for future issues.

This year, an impressive number of Capstone projects – fifteen in total – were awarded 'Thesis of Distinction' by the AUC Capstone Awards Committee. The three papers published here were selected from this group

and represent undergraduate scholarship from each of the three majors at AUC: Humanities, Sciences, and Social Sciences. They are also among those projects selected by the Committee for 'Thesis of Highest Distinction', an award reserved for the most exceptional student work.

The papers in this issue may be award-winning, but more importantly they are representative of what distinguishes our students' Capstone projects across the board. They demonstrate creativity, depth of enquiry, and originality of thought.

Following the inaugural issue, it is our plan to bring students into the editorial and review process, making the Journal, like AUC itself, a place of collaboration within the academic community, joining together the diverse interests and expertise of our students and faculty.

Prof. Dr. Marijk van der Wende,
Dean

Dr. Rebecca Lindner,
Head of Studies, Humanities

CONTENTS

- 6** The Persistence of the Winner's and Loser's Curse in a One-Player Decision Game
Karlijn Hoyer
- 34** The Limits of Friendship
Alex M. Verbeek
- 60** The Effects of Serotonin in the Nucleus Accumbens Shell on Glucose Metabolism in Rats
Kayleigh van Megen

The Persistence of the Winner's and Loser's Curse in a One-Player Decision Game

Karlijn Hoyer



ABSTRACT

Experimental evidence supports both the existence and the persistence of the winners and losers curse, suggesting that agents are not only irrational, but also do not succeed in learning to be rational. This thesis investigates whether the losers curse is more persistent than the winners curse in a one-player decision game based on Bazerman and Samuelsons (1985) buy-a-firm game, using a simplified version of Feltovichs (2006) reinforcement learning model. The following two hypotheses are tested: 1) slow or no convergence to optimal play occurs in both the winners and losers curse situation and 2) convergence to optimal play in the losers curse situation takes place faster than in the winners curse situation. Given the prevalence of auctions in business today, it is extremely important for decision-makers to fully understand the the winners and losers curse, as over- and underbidding can have disastrous consequences on individual budgets and businesses. In support of hypothesis 1, I find an extremely slow convergence to optimal play in both contexts. The first 50 rounds suggest that hypothesis two holds, albeit not statistically significant, and the long-term behaviour (20,000 rounds) suggests the opposite. It is left to further research to determine which effect dominates empirically.

1. INTRODUCTION

Economics often assumes that all agents' behaviour is rational, i.e. that their choices are consistent with well-defined preferences. However, many empirical results are difficult to rationalize without turning to implausible assumptions. One of these anomalies, which often occurs in laboratory experiments using markets with asymmetric information and common value auctions, is called the winner's curse (overbidding). The opposite bias is called the loser's curse (underbidding).

The name winner's curse originates from Capen, Clapp and Campbell (1971). The winner of the auction can be said to be 'cursed' in two ways: 1) the winning bid exceeds the value of the object purchased, leading to negative profits, or 2) the value of the object purchased is less than the expected value, leading to lower profits than expected. In the first situation the winner loses, in the latter the winner is disappointed (Thaler, 1988). In reality, firm managers

who consistently fall prey to the winner's curse would go bankrupt. The winner's curse is defined as 'the systematic failure of bidders to account for the adverse selection process whereby winning bids produce below normal or even negative profits' (Kagel & Roth, 1995).

There might also be a second factor that can induce overbidding: the 'thrill of winning'. It is possible that overbidding is due to a 'utility of winning' instead of being the result of an agent failing to anticipate on the adverse selection process. In order to distinguish between the two situations in which the winner's curse can occur, the winner's curse must be neutralised by identifying this opposing bias. For this reason Holt and Sherman (1994) consider the loser's curse. This new situation allows the range of bidding values to be larger than the range of values that the seller might accept. In reality, firm managers who fall prey to the loser's curse would make purchases too infrequently, which is also detrimental to business.

The loser's curse is defined as 'a situation in which a failure to anticipate the informational content of a bid's acceptance will cause one to bid below the optimal bid' (Holt & Sherman, 1994).

Holt and Sherman conclude that both the loser's curse and the winner's curse can be a result of naïve bidding, suggesting that if all bidders were rational, the winner's and the loser's curse would not occur. However, acting rationally in a situation with asymmetric information can be extremely difficult, as this requires agents to distinguish between the expected value of the object for sale (conditioned on the prior information available to them) and the expected value of winning the auction. People often fail to take the asymmetric information that is built into the problem into consideration in their analysis of the problem. According to Thaler (1988), traditional economists usually respond to examples like this by hypothesising that people can only be fooled once or twice and that with growing experience they will figure out the trap.

Interestingly, experimental evidence supports both the existence and the persistence of the winner's curse, suggesting that agents are not only irrational, but also do not succeed in 'learning' to be rational. Learning is defined as an increase in performance level over time: the fraction of 1) optimal bids increases, 2) the fraction of winner's curse or loser's curse bids decreases and 3) the fraction of dispersed bids decreases (Casari, Jackson, & Zhang, 2009).

Few experiments show learning. This learning is never easy nor fast. Outside of the economic laboratory, many studies also claim to have found real world evidence of the winner's curse in market contexts. Therefore, more insight into the mechanisms underlying these biases can be extremely useful. In contrast to the winner's curse research, only a limited amount of experimental research has been conducted on the existence and especially on the persistence of the loser's curse. Furthermore, the difference between learning in winner's and loser's curse situations has never been examined.

This thesis investigates whether the loser's curse is more persistent than the winner's curse in a one-player decision game based on Bazerman and Samuelson's (1985) buy-a-firm problem. I hypothesize the following:

Hypothesis 1: Slow or no convergence to optimal play occurs in both the winner's and loser's curse

situations. Slow learning in the winner's curse situation is in accordance with the findings from the literature. In addition, I expect that the convergence of the loser's curse also takes place slowly, as the initial decision problem is similar.

Hypothesis 2: Convergence to optimal play in the loser's curse situation takes place significantly faster than in the winner's curse situation. Feltovich's (2006) model of reinforcement learning suggests that loser's curse experiments might converge faster than winner's curse experiments. In accordance with this model, I expect that people learn to overcome the loser's curse more quickly than they learn to overcome the winner's curse.

In an effort to understand this, I will use a simplified reinforcement learning model based on Feltovich's (2006) model in order to investigate the possibility and rate of convergence towards optimal play in both winner's and loser's curse contexts, and to test both hypotheses.

The set-up of my thesis is as follows. Chapter 2 reviews the relevant literature and familiarizes the reader with the most important concepts. Chapter 3 introduces the simplified model and simulates the behaviour of one player participating in a one-player decision game over several round lengths. Chapter 4 reruns the same simulations for multiple players participating and investigates the convergence rates of both curses. Chapter 5 concludes, discusses the implications of my results for businesses, and gives suggestions for further research.

2. LITERATURE REVIEW

2.1. INTRODUCTION

The phenomena of the winner's and loser's curse have been the subject of interest of many research studies related to bidding behaviour and errors in bidding decisions. This chapter first deepens the understanding of the main concepts involved (Section 2.2) and reviews experimental and field evidence in support of the existence and persistence of the winner's curse, and to a lesser extent the loser's curse, both in- and outside of the economics laboratory (Section 2.3). In addition, this chapter discusses Holt and Sherman's (1994) static model of bidding and biases in order to gain insight into the difference between naïve and rational bidding behaviours. Finally, I will discuss Feltovich's (2006) reinforcement learning model that captures the learning dynamics of the persistence of both curses (Section 2.4).

2.2. THE CONCEPTS

This subsection reviews the most important terms used throughout this thesis. I will first define the winner's (Section 2.2.1) and loser's curse (Section 2.2.2), and discuss the economic intuition behind both curses. Section 2.2.3 introduces the buy-a-firm game. This game forms the basis of various laboratory experiments in this field (Section 2.3.1) and also of Feltovich's (Section 2.4) and my reinforcement learning models (Chapters 3 and 4).

2.2.1. THE WINNER'S CURSE

The name winner's curse originates from Capen, Clapp, and Campbell (1971). This concept is mostly discussed in the context of oil companies that want to purchase drilling rights to a specific parcel of land, and therefore enter a common value auction. In a common value auction, all bidders have different private information about the value of the object for sale, but the actual value of this object is the same for all of them. In this first price auction, the bidder who submitted the highest bid 'wins' and pays a price equal to the amount of the bid. Before bidding, all companies let their experts estimate the value (E) of the drilling rights. Assuming that these estimates are unbiased, the mean of all estimates will be equal to the common value (\bar{E} in Figure 1). However, estimating the total amount of oil in a given parcel is

extremely difficult. Experts' estimates will thus vary from far too high to far too low. A bid (B) will equal the amount estimated minus a discount factor D , as firms want to make profits, $D = E - B$.

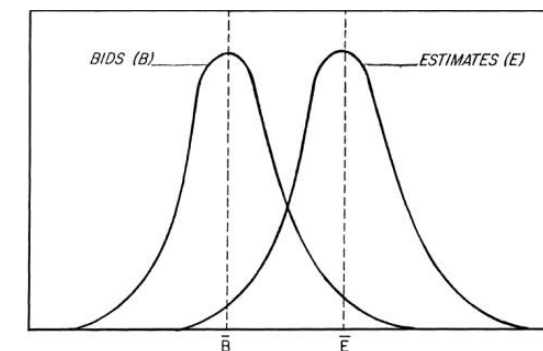


Figure 1: Graphic Illustration of the winner's curse (Bazerman & Samuelson, 1983). The winning bid exceeds the common value.

The firm that provides the highest estimate is more likely to bid higher and win the auction than a firm with a lower estimate. Thus, chances are that the winner of the auction is actually a loser, as the winner will receive negative or lower profits than expected (Thaler, 1988). In other words, a buyer should realize that a bid based on an overestimate is much more likely to be accepted, resulting in a win which the buyer will regret afterwards (Holt & Sherman, 1994).

If all bidders are rational, the winner's curse will not occur. However, as explained above, acting rationally in such an auction can be extremely difficult, as this requires agents to distinguish between the expected value of an object for sale (conditioned on the prior information available to them) and the expected value of winning the auction. These two factors, working in opposite directions, need to be considered. On the one hand, this means that an agent needs to bid more aggressively in order to win if the number of bidders increases. On the other hand, an increase of the number of bidders also increases the chance that the winner has overestimated the value of the object for sale, suggesting a less aggressive bidding strategy (Thaler, 1988). People often fail to take the asymmetric information that is built into the problem into consideration in their analysis of the problem.

2.2.2. HOLT AND SHERMAN'S LOSER'S CURSE

There is also a second factor that can induce overbidding: the thrill of winning. It is possible that

overbidding is due to a 'utility of winning' instead of being the result of an agent failing to anticipate on the informational content of a bid's acceptance. For that reason, Holt and Sherman (1994) consider the loser's curse. In order to distinguish between the two situations in which the winner's curse can occur, the winner's curse must be neutralised by identifying this opposing bias. In this new situation, an agent who fails to anticipate on the informational content of a bid's acceptance will bid below the optimal bid: underbidding.

The intuition behind the possibility of underbidding is that 'an increase in the bid at the margin from B to $B + \Delta$, will only matter if the seller value is between B and $B + \Delta$ [i.e., if the seller value is high relative to the range of seller values below B] (Holt & Sherman, 1994, p. 649). This is shown in Figure 2 where v is uniformly distributed on $[0.5, 1]$ with a lower bound greater than zero. This allows the range of bidding values to be larger than the range of values that the seller might accept. For example, if $B = 0.7$, then v is uniformly distributed on $[0.5, 0.7]$ if the bid is accepted (dotted line). The value for the buyer $(1.5v)$, (dashed line) is now uniformly distributed on $[0.75, 1.05]$. Therefore, the buyer will always earn a profit as $0.7 < 0.75$.

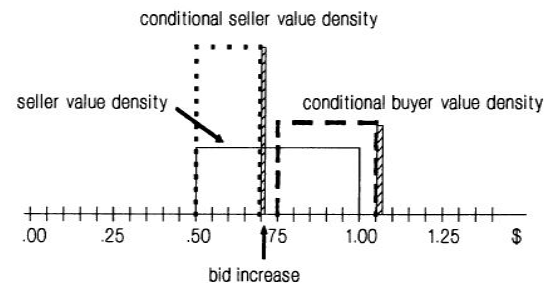


Figure 2: The effect of an increase in the bid on value distribution (Holt & Sherman, 1994).

Increasing the bid from 0.7 to $0.7 + \Delta$, will only be relevant if the seller value is between 0.7 and $0.7 + \Delta$, which is higher than the seller values on $[0.5, 0.7]$. In this case, a bid increase at the margin picks up relatively high value units (shaded lines). If a naïve bidder does not adjust for this, the bidder will bid too low and will win too infrequently.

2.2.3. THE BUY-A-FIRM PROBLEM

In order to experimentally investigate the winner's curse, Bazerman and Samuelson (1985) conducted several experiments using the 'buy-a-firm' problem (Thaler, 1988). In this problem, company A (acquirer) wants to buy company T (target), and company T accepts every bid which is greater than or equal to the value of the company under its own management. Assume that the value v of T is uniformly distributed on $[X, X + R]$. Company T's advantage is that it can observe its value v . Company A knows that its value for company T equals Mv with $M > 1$. Company A makes a single bid (Holt & Sherman, 1994). Figure 3 illustrates the case in which $X = 0$, and $B = 0.5$ and $M = 1.5$. The expected value of a bid would equal $E(v | B \geq v) = \frac{B}{2}$. For company A, company T would in this case be worth only $1.5E(v | B \geq v) = \frac{3B}{4}$. This means that every positive bid yields an expected loss to the bidder, since $\frac{3B}{4} < B$ (Selten, Abbink, & Cox, 2005).

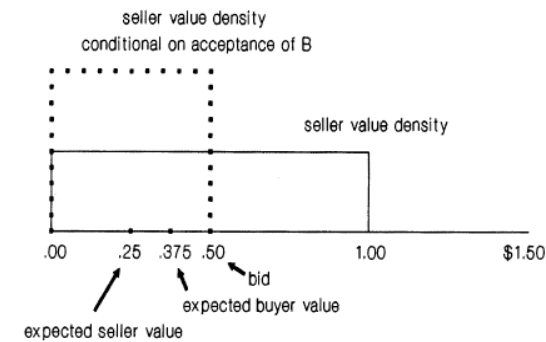


Figure 3: Picture illustrating the winner's curse in the buy-a-firm problem (Holt & Sherman, 1994).

Holt and Sherman (1994) also developed a more general static model of bidding and biases in which parameter value X can be varied in such a way that a winner's curse ($X = 0$), a loser's curse ($X = 10$), and no curse ($X = 5$) situations can be created, as described below. In the no curse situation, a naïve bidder would choose a bid that just happens to maximize expected earnings. Their model gives an insight in the differences between naïve and rational bidding behaviours.

In their model, $\frac{B-X}{R}$ is the acceptance probability of the bid. Taking the current owner's decision rule into account, the expected value of the firm is $X + \frac{B-X}{2}$. A risk-neutral bidder's optimal bid maximizes the earning integral: $\int_X^B (Mv - B) dv$. Expressing this as the product of the acceptance probability and the expected

earnings conditional on acceptance yields the rational objective:

$$\left(\frac{B-X}{R}\right)\left[M\left(X + \frac{B-X}{2}\right) - B\right].$$

In contrast, a naïve bidder does not consider the current owner's decision rule. This agent's expected value of v is $X + \frac{B}{2}$ and his own value is thus $M\left(X + \frac{B}{2}\right)$. This gives the naïve objective:

$$\left(\frac{B-X}{R}\right)\left[M\left(X + \frac{B}{2}\right) - B\right].$$

This results in the following first order conditions when differentiating with respect to B . For rational bidding:

$$(1) \quad M\left(X + \frac{B-X}{2}\right) - B + (B-X)\frac{M}{2} - (B-X) = 0.$$

For naïve bidding:

$$(2) \quad M\left(X + \frac{B}{2}\right) - B - (B-X) = 0.$$

The first terms in equation (1) and (2) represent the buyer's expected earning conditioned on acceptance ($Mv - B$), as described above. The second term in equation (1), $(B-X)\frac{M}{2}$, represents the rational observation that B is the upper limit of the seller's value if the bid is accepted. This term is proportional to the probability of making the purchase. The intuition about the loser's curse described earlier holds with this equation. Rewriting equation (1) and (2) in terms of B_R (rational bid) and B_N (naïve bid) gives:

$$B_R = \left(\frac{1}{2-M}\right)X;$$

$$B_N = \left(\frac{M+1}{2}\right)X + \left(\frac{M}{4}\right)R.$$

X and R can be varied to create or balance the tendencies to overbid and/or underbid. $X = 0$ represents a winner's curse environment, $X = 5$ a no curse environment and $X = 10$ a loser's curse environment.

2.3. EXISTENCE AND PERSISTENCE

This subsection reviews experimental and field evidence in support of the existence and persistence of the winner's curse, and to a lesser extent the loser's curse, both inside (Section 2.3.1) and outside (Section 2.3.2) of the economics laboratory.

2.3.1. EXPERIMENTAL EVIDENCE

There has been ample experimental evidence for the existence and persistence of (especially) the winner's curse, in particular from laboratory experiments. Bazerman and Samuelson (1983) conducted a jar of coins experiment with 419 MBA students in a first price sealed bid auction format. Each jar of coins had a value of \$8.00, which was unknown to the students. The demonstrator circulated the closed jar among the students, so that every individual student could

examine the jar. In total 48 auctions were conducted, 4 in each of the 12 microeconomics classes. The group size varied per class and subjects were familiar with the group size for each auction beforehand. Subjects were also told that the highest bidder would win the value of the jar, e.g. if the highest bid was \$6.00, the winner would receive \$8.00-\$6.00 (thus \$2.00). Furthermore, a \$2.00 prize was given to the closest estimate of the true value in each auction, in order to promote the best possible estimates. No feedback was provided until the end of the experiment. They found the following results: the mean estimate of the value was \$5.13 (\$2.87 below the real value) and the mean winning bid was \$10.01, leading to an average loss of \$2.01 for the winning bidder. This clearly provides evidence in favour of the existence of a winner's curse, despite the significant underestimation of the value.

Bazerman and Samuelson (1985) have repeated similar experiments in different contexts, e.g. the buy-a-firm problem (Thaler, 1988). Their experiment was run in two conditions, one with monetary incentives and one without. They found the following results: in both conditions over 90% (92% in the first and 91% in the second) of the subjects made positive bids. The average bid was slightly lower in the monetary incentive condition. The majority of the bids were between \$0.50 and \$0.75, which clearly is much larger than the optimal bid of \$0. Participants almost always made a loss. Over the years, many experiments have been conducted, replicating the result of the above explained fundamental experiments by Bazerman and Samuelson. Even when placed in a different context and when parameters are varied, the results support the existence of the winner's curse.

In 1987 Weiner, Bazerman, and Carroll (WBC) have investigated the hypothesis that people learn to avoid the winner's curse by using a modified version of the buy-a-firm problem. All of WBC's subjects individually repeated the experiment 20 times. The value of company T was independent in each session and uniformly distributed on $[0, 1]$. One group received monetary incentives, one group did not. Subjects were told that they were participating in a study on how people make investment decisions under uncertainty and that the actual acquisition of the company was a neutral event, meaning that one's performance was not measured by whether they 'won' company T or not. The monetary incentive group was also told that their pay-offs depended on their performance. After

each trial, the subjects received feedback, including the real value of company T, whether their bid was accepted by company T, how much money they made or lost and how much money they had spent (their initial endowment minus their (negative) profits). WBC found that out of the 69 subjects, only 5 learned to bid \$0.10 or less after on average 8 trials. Among the others, there was no sign of any learning. They also found no significant differences between the group that received monetary incentives and the group that did not. This suggests that people may be able to learn to avoid the winner's curse, but very slowly (Thaler, 1988). Horking (1990) also observed little learning in an identical experiment. Similarly, Ball, Bazerman and Caroll report that fewer than 10% learned to bid \$0 and the mean of the others was \$0.50 (Ball, Bazerman, & Carroll, 1991).

Kagel and Levin (1986) have used the same set-up with more bidders trying to acquire company T at the same time. They observe losses when the group of bidders is large (6-7 people). These results have been replicated by Kagel, Levin, and Harstad (1987) using a second price auction method, in which the highest bidder wins and pays the second highest bid. The profits gained by the winner of a small group of other bidders (3-5 bidders) are about 50-65% of the sum the player could have received if the player would have played rationally. In 1987 Dyer, Kagel, and Levin (DKL) reported on low price auctions, in which the winner is the one who submits the lowest unique bid. They found losses in both small and larger groups of bidders (Thaler, 1988). Interestingly, the DKL experiment was conducted with firm managers, who would in reality go bankrupt if they would consistently fall prey to the winner's curse. DKL state that perhaps managers learned situation-specific rules of thumb rather than theory. This shows that the statement commonly used in defence of rationality, namely that subjects are simply students and that real world experts would not make these kinds of mistakes, is wrong (Thaler, 1988).

A more recent experiment by Casari, Jackson, and Zhang (CJZ, 2009), again using the same 'buy-a-firm' problem, reports on decision-making procedures of individuals versus groups when facing a winner's curse. CJZ manipulated the decision making process by using three treatments: individual decision making (individual), individual decision making when observing the bids of two other people (signal) and

group decision making (group). They conducted eight experimental sessions with 15 participants, thus a total of 120 people. After each trial, the participants received the same feedback as in WBC's experiment. Participants were asked to write down their bids and a confidence level associated with their bids. They were paid based on their performance. CJZ found the following two results 1) there was no significant learning over time in the individual treatment, and 2) there was significant learning in the group treatment. Interestingly, they also found that individuals that can observe the bids placed by two other participants (signal) did not perform significantly better in comparison to the individual treatment. Earlier experiments, e.g. by Chua and Luk (2005), support the result that groups are superior to individuals in making optimal bidding choices, leading to a lessening of the winner's curse.

The evidence presented above shows that avoiding the winner's curse is not easy and that even subjects that have the opportunity to learn, fail to solve the problem. Also, groups are more likely to reach an optimal bidding strategy than individual bidders. Group decision-making goes beyond the scope of this thesis, as I will focus solely on one-player decision games.

2.3.2. FIELD EVIDENCE

Outside of the economic laboratory, many studies claim to have found real world evidence of the winner's curse in market contexts. Besides the oil and gas drilling rights discussed by Clapp and Campbell, Dessauer (1981) reports that in the field of book publishing, most auctioned books had a perceived value higher than their real value, meaning that people would bid more for a book than it was actually worth. More evidence comes from Cassing and Douglas (1980), who report that free agents in baseball are overpaid. Roll (1986) applies the concept of the winner's curse to corporate takeovers, to discover why firms are willing to pay substantial premiums above the market price to acquire another firm. His evidence suggests that there is little or no gain for the buying firm. This is quite similar to the 'buy-a-firm' results reported from the laboratory. Thaler (1988) also reports on evidence from offshore oil and gas leases, researched by various authors including Capen, Clapp, and Campbell (1971) and Hendricks, Porter, and Boudreau (1987).

Another interesting example occurred during the third generation (3G) UMTS spectrum auction by the UK government. In 2000, the UK telecommunications companies paid £22.47 billion for five 3G mobile telephone licenses. This greatly exceeded both the government's initial expectations and the average price paid per head of population (£107.20) in other countries during similar auctions, e.g. in Germany (£93.10), Italy (£35.20) and Switzerland (£2.60-£24.70). Looking at the stock markets after the auctions suggests that the UK companies have been punished for paying too much. Post-auction, the UK telecommunications companies appear to have fallen for the winner's curse (Cable, Henley, & Holland, 2002).

More recently, the National Broadcasting Company (NBC) lost \$223 million on the 2010 Winter Olympics, most of it due to the expenses associated with acquiring the rights to broadcast the Vancouver Games on television and internet. NBC paid over 30% more than they paid in 2006. These losses were certainly not caused by an extremely low viewership that resulted in poor revenues unable to cover the costs, as officials say that the 2010 Winter Olympics had 14% better ratings than the 2006 Winter Olympics in Turin, Italy. Apparently, NBC was the network that had the rosiest outlook on the revenue that the Olympics would generate. Did NBC fall prey to the winner's curse (Beggs, 2010)?

2.4. FELTOVICH'S REINFORCEMENT LEARNING

Reinforcement models assume that players choose which strategy they play with a certain probability. The Law of Effect formulated by Thorndike (1898) states that choices that have led to better outcomes in the past become more likely in the future, meaning that strategies that have done well in the past have a higher probability in the future and vice versa. The rules for this learning model are typically formulated as follows: At time $t = 1$ every player n has an initial propensity $q_{nk}(t)$ to play each of his k strategies. The strategies are updated by adding the payoff x to the original propensities: $q_{nk}(t+1) = q_{nk}(t) + x$ for the strategy played and $q_{nk}(t+1) = q_{nk}(t)$ for all other strategies. The probability that player n plays strategy k at time t is given by $p_{nk}(t) = q_{nk}(t) / \sum_j q_{nj}(t)$ (Salmon, 2001). An agent's strength of propensities is given by $Q_n^t = \sum_{j=0}^9 q_n^t(j)$. In this model, agents' behaviour changes

very little when a bid is rejected and therefore, learning takes place very slowly and even repeated mistakes are likely. In addition, it might sometimes happen that a bad action yields a high pay-off, making a bad action more likely in the future: learning is noisy.

Feltovich (2006) argues that the learning method in these situations is learning via reinforcement. Feltovich's starting point is a naive model similar to Holt and Sherman's model. He uses the same rules with a slightly different notation: λv is the buyer value and the buyer knows that $v = \{x_0, x_0 + 1, x_0 + 2, x_0 + 3, \dots, x_0 + 9\}$.

Feltovich based his one-player decision learning model on the adjustable-reference-point reinforcement learning model of Erav and Roth (1996), allowing for both (persistent) suboptimal decisions and learning. The model has five parameters: $\epsilon, \delta, w^+, w^-$ and Q_n^0 ; their initial values are summarised in Table 1. Strategy $k \in \{0, 1, 2, \dots, 9\}$ corresponds to a bid $x_0 + k$. The pay-off received in round t depends on the strategy the agent chose and on the seller's realized values. The pay-off is given as:

$$\prod_n^t(k, v) = \begin{cases} \lambda v - (k + x_0), & k + x_0 \geq v \\ 0, & k + x_0 < v \end{cases}$$

Bidders update their propensities based on their pay-off by comparing this pay-off to a reference point ρ_n^t . This reference point can vary over time. The initial reference point is set equal to the minimum possible pay-off, $(\lambda - 1)x_0 - 9$, and is updated each round based on the pay-off received as given in the following equation, in which w^+ represents outcomes that are better than expected and w^- represents outcomes that are worse than expected:

$$\rho_n^{t+1} = \begin{cases} (1 - w^+) (\rho_n^t) + (w^+) \prod_n^t(k, v), & \prod_n^t(k, v) \geq \rho_n^t \\ (1 - w^-) (\rho_n^t) + (w^-) \prod_n^t(k, v), & \prod_n^t(k, v) < \rho_n^t \end{cases}$$

Most reinforcement originates from the action played. However, almost similar actions are also slightly reinforced. Reinforcement goes as follows:

$$q_n^{t+1}(k) = (1 - \delta) q_n^t(k) + (1 - \epsilon) \left[\prod_n^t(k, v) - \rho_n^t \right]$$

$$q_n^{t+1}(j) = (1 - \delta) q_n^t(j) + \frac{\epsilon}{M} \left[\prod_n^t(k, v) - \rho_n^t \right] \quad \text{for } |j - k| = 1$$

$$q_n^{t+1}(j) = (1 - \delta) q_n^t(j) \quad \text{for } |j - k| > 1$$

M is the number of strategies that are similar to k , $\epsilon \in [0, 1]$ represents the amount of experimentation between strategies, and $\delta \in [0, 1]$ reflects 'forgetting'.

Feltovich (2006) runs a simulation and concludes that average bids tend to move in the direction of optimal choices over time, but extremely slowly so. Reducing λ leads to faster convergence to optimal play.

Also, he finds evidence that loser's curse experiments might converge faster.

Parameter	Q_0	ϵ	δ	w^-	w^+
Value	$(\lambda-1)x$	0.1	0.001	0.01	0.02

Table 1: Learning model parameter values (Feltovich, 2006).

2.5. CONCLUSION

This chapter reviewed experimental and field evidence indicating that the winner's curse exists and that people often fail to take the information asymmetry that is built into the problem into consideration. By considering the opposite bias, Holt and Sherman (1994) conclude that both the loser's and the winner's curse can result from naïve bidding. Their model provides an insight into the difference between naïve and rational bidding behaviours. On top of that, experimental evidence suggests that the winner's curse persists even if agents are given learning opportunities by adding feedback to the 'buy-a-firm' problem., e.g. by Weiner, Bazerman, and Carroll (1987), Kagel, Levin, and Harstad (1987). and Casari, Jackson, and Zhang (2009). Furthermore, this chapter reviewed Feltovich's (2006) reinforcement learning model. He found a slow convergence towards the optimum and suggests, but does not prove, that the loser's curse converges faster. The model that I will develop in the next chapter is a simplified version of his reinforcement learning model.

3. SIMULATIONS OF ONE PLAYER

3.1. INTRODUCTION

This chapter describes the results from my simulations, based on a simplified version of Feltovich's (2006) reinforcement learning model discussed in the previous chapter. The model captures the essential decision making features in a one-player decision game based on the buy-a-firm problem over multiple time periods. The simulations in this chapter cover the results of one player participating in a one-player decision game, in contrast to Chapter 4, where the results of multiple players participating in the same game are averaged. Section 3.2 explains the model and its features in detail. Section 3.3 reviews the results of my simulations and compares these to the original results by Feltovich. Finally, Section 3.4 gives conclusions.

3.2. REINFORCEMENT LEARNING MODEL

This section will explain the features of the decision problem underlying my reinforcement learning model and explain how the learning dynamics take place. Furthermore, a subsection is devoted to the differences between Feltovich's model and my model.

3.2.1. THE DECISION PROBLEM

The idea of the reinforcement learning model used throughout this chapter is as follows. The value of firm T (target) under its own management is given by v . T accepts every bid $b \geq v$ by firm A (acquirer) and rejects $b < v$. The acceptance of the bid involves no strategic decision making and the game is therefore reduced to a one-player decision problem. Every round starting at $t = 0$, v is randomly drawn from a uniform distribution on the set $\{x, x+1, x+2, x+3, \dots, x+9\}$ with $x \in \{0, 10\}$. Company T (the seller) knows v , whilst firm A (acquirer, buyer) is only aware of the value distribution of v . For firm A, firm T is worth λv with $1 \leq \lambda \leq 2$. Firm A can pick any bid $b = x + k$, where $k \in \{0, 1, 2, \dots, 9\}$. The outcomes of the game are dependent on the value of v and b .

In accordance with the literature, setting $X = 0$ in the simulations represents the winner's curse situation and $X = 10$ represents the loser's curse situation. The optimal bid in the winner's curse situation is 0. The optimal bid in the loser's curse situation corresponds to $19 (= x + 9)$. Obviously, bidding $b > x + 9$ is dominated by bidding $b = x + 9$, as

this will lead to a higher profit, since the maximum value is $v = x + 9$ and every bid $b \geq v$ is accepted. A bid $b < v$ is never accepted and pays zero with certainty. This bid is dominated by a bid of exactly $b = x$, since this pays zero unless it is accepted, which always yields a positive pay-off. This means that every bid outside of the value distribution of v is dominated by some bid inside the value distribution. Using the equation for the rational bid defined by Holt and Sherman (see Chapter 2), one can easily check that the optimal bids mentioned above hold:

$$B_r = \left(\frac{1}{2 + \lambda} \right) x$$

For the winner's curse

$$x = 0 \quad B_r(x = 0, \lambda = 1.5) = \left(\frac{1}{2 - 1.5} \right) 0 = 0.$$

For the loser's curse $X = 10$,

$$B_r(x = 10, \lambda = 1.5) = \left(\frac{1}{2 - 1.5} \right) 10 = 2 * 10 = 20, \text{ and since } 20 \text{ is outside of the value distribution, the optimal bid in this case equals } 19.$$

3.2.2. THE LEARNING DYNAMICS

A player's learning dynamics are determined by the propensities of the possible strategies, which are updated over time. At time $t = 0$, every strategy is equally likely to be played, as the initial propensities of all strategies q_k are equal. The probability that the player plays strategy k at time t is given by $p_k(t) = q_k(t) / \sum_j q_j(t)$. After every round, the player compares the pay-off that he received for playing strategy k at time t to the average pay-off that he received throughout the game. If this difference is positive, then the propensity corresponding to the strategy played will be positively reinforced, which makes it more likely that this strategy will be played in the future. If this difference is negative, then the propensity will be negatively reinforced and thus this strategy becomes less likely. The reference value can be seen as the amount of money that the player expects to win based on previous rounds.

A player's profit is determined by:

$$\prod^t(k, v) = \begin{cases} \lambda v - b, & b \geq v \\ 0, & b < v \end{cases}$$

This equation shows that the bidder receives zero profit when his bid is smaller than the value of firm T, as in this case his bid is not accepted. When his bid equals or exceeds the value of firm T, profits are determined by the difference between buyer value and the bid.

Bidders update their propensities as follows:

$$q^{t+1}(k) = q^t(k) + \left[\prod^t(k, v) - \rho^t \right]_{j \neq k}$$

$$q^{t+1}(j) = q^t(j),$$

The propensities of the strategies that were not played in a specific round are not updated and remain the same. The propensity of the strategy played is updated by adding the difference between the obtained profits and the reference value of the pay-offs. The reference value of the pay-offs represents the average profits obtained up to that round and is given by equation:

$$\rho^{t+1} = \rho^t - \left(\rho^t - \prod^t(k, v) \right) / t$$

As mentioned before, every round, v is randomly drawn from a uniform distribution on the set $\{x, x+1, x+2, x+3, \dots, x+9\}$ with $x \in \{0, 10\}$. At $t = 0$, k is randomly drawn from $k = \{0, 1, 2, \dots, 9\}$. After the first round, the strategy is randomly chosen from a distribution taking both the strategies and their weights into account. All initial propensity values are fixed at $q_k = 5$. If propensities are reinforced below 0.001, they are set back to 0.001 automatically. The initial reference value of pay-offs is 0. The amount of strategies and value of λ can be varied in the simulation, but will be fixed at 10 (from 0 to 9) and 1.5, respectively, for the time being. The numbers of cycles I will consider when evaluating the long-term behaviour of both curses are $t = 50$, $t = 500$ and $t = 20,000$. The maximum number of rounds in which participants in an economic laboratory may participate is approximately equal to 50 and therefore the results of the simulation with $t = 50$ can easily be compared to participants' behaviour in such an experiment. I use $t = 500$ for comparison reasons, as this is the amount of cycles that Feltovich uses. $t = 20,000$ is used to investigate whether the long-term behaviour settles at a specific average strategy. The simulation is programmed in JavaScript.

3.2.3. DIFFERENCES WITH FELTOVICH'S MODEL

The reinforcement model that I use is considerably simplified in comparison to Feltovich's original model: a different reference value is used, different initial propensities are used, neighbouring strategies are not reinforced and various parameters are set to zero. The changes are summarised in Table 2 and Table 3.

Firstly, the model used in this chapter uses the average pay-off as reference value. In comparison, Feltovich uses $(\lambda - 1)x - 9$, the minimal possible pay-

off, as initial reference value, yielding -9 in the $x=0$ case and -4 in the $x=10$ case. I find this counter-intuitive, since few people will enter an auction if they expect to make negative profits in the first place. In order to make the simulations more realistic, I have chosen to set the reference value equal to the average payoff. "Thinking in averages" might be problematic in the long term and therefore, not completely representative to the real world. The model can, however, be used to compare experimental results from a 50 period session in which this assumption is still valid (see Section 5.3). For simplification, I assume that people are capable of calculating averages over longer time periods.

Secondly, in order to ensure that the propensities do not become negative, as this would yield negative probabilities, I fix all initial propensities to 5, and propensities that are reinforced below 0.001 are set back to 0.001 automatically. Feltovich uses $Q = (\lambda - 1)x$ as initial value for the summation of all propensities, which is roughly the average magnitude of payoffs. As propensities are equal in the first round, this would yield 0 in $x=0$ and 0.5 in $x=10$ for all ten propensity values. In this case, a negative reinforcement at the beginning of the simulation would result in negative probabilities and unrepresentative results. Setting all propensities to 5 in the beginning is not a problem, since the most important part is that all propensities are equal. I could also have chosen any other value greater than zero.

Parameter	Q_0	ϵ	δ	w^-	w^+	ρ^o	q_{\min}^k
Value (Feltovich)	$(\lambda-1)x$ or $3(\lambda-1)x$	0.1	0.001	0.01	0.02	$(\lambda-1)x-9$	n/a
Value	50	0	0	n/a	n/a	0	0.001

Table 2: Differences in learning model parameter values.

Thirdly, I made various simplifying adjustments to the model, namely that the neighbouring strategies are not reinforced and that the parameters ϵ – experimentation value – and δ – gradual forgetting parameter – are set to zero. Since the amount of strategies that can be played is not too large, people are easily able to distinguish between strategies and thus, not reinforcing the neighbouring strategies of the strategy played is a realistic change to the model. Throughout the simulation, the number of strategies is fixed at 10, which still makes it manageable for participants in an experiment to distinguish between

strategies, and therefore only reinforce the strategy that was played. For simplification, I also assume δ , the gradual forgetting parameter, to be zero. All parameters are reintroduced to the model in Section 4.3 in order to isolate their individual effects on the learning outcomes.

Furthermore, I will not consider the situation with $x=5$, also known as the no curse situation, as this goes beyond the scope of this thesis. Feltovich devotes more attention to this no curse example.

Model	Propensities
Feltovich	$q_n^{t+1}(k) = (1-\delta)q_n^t(k) + (1-\epsilon) \left[\prod_n^t(k,v) - \rho_n^t \right]$
	$q_n^{t+1}(j) = (1-\delta)q_n^t(j) + \frac{\epsilon}{M} \left[\prod_n^t(k,v) - \rho_n^t \right], \text{ for } j-k =1$
	$q_n^{t+1}(j) = (1-\delta)q_n^t(j), \text{ for } j-k >1$
	Reference point
	$\rho_n^{t+1} = \begin{cases} (1-w^+) (\rho_n^t) + (w^+) \prod_n^t(k,v), \text{ for } \prod_n^t(k,v) \geq \rho_n^t \\ (1-w^-) (\rho_n^t) + (w^-) \prod_n^t(k,v), \text{ for } \prod_n^t(k,v) < \rho_n^t \end{cases}$
My Model	Propensities
	$q^{t+1}(k) = q^t(k) + \left[\prod^t(k,v) - \rho^t \right]$
	$q^{t+1}(j) = q^t(j) \text{ for } j \neq k$
	Reference point
	$\rho^{t+1} = \rho^t - \left(\rho^t - \prod^t(k,v) \right) / t$

Table 3: Differences in reinforcement methods.

3.3. RESULTS

3.3.1. THE WINNER'S CURSE SITUATION

Throughout this set of simulations, x is fixed at zero: the winner's curse situation. As we have seen before, the optimal bid in this situation is zero. Learning would therefore require convergence towards zero over time: an increase in optimal bids, a decrease in non-optimal bids, and a decrease in dispersed bids.

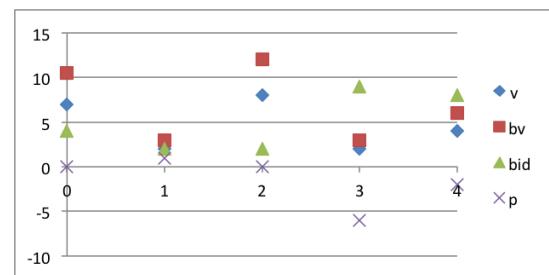


Figure 4: Simulation of the winner's curse situation over 5 time periods indicating the value (v), the buyer value (bv), the bid (b) and the payoff (p) in euros.

Figure 4 gives a first intuition of what is happening in this situation, with the amount of Euros on the vertical axis and time on the horizontal axis. The pay-off is determined by the difference between the buyer value (bv), which equals 1.5 times the value (v), and the bid (b), which equals strategy (k) plus the lower bound of the distribution (x) of the value. Value (v) is randomly picked between 0-9 in each round. The figure shows that if a bid is not accepted, the pay-off is zero, as in round $t=0$ and $t=2$. Furthermore, when the bid is accepted by the seller, we see that it yields a slightly positive pay-off at $t=1$ but a negative pay-off at $t=3$ and $t=4$. When running the simulation over multiple time periods, the number of bids resulting in a negative pay-off increases.

Learning can be determined by plotting the average strategy over time, as is done in Figure 5.

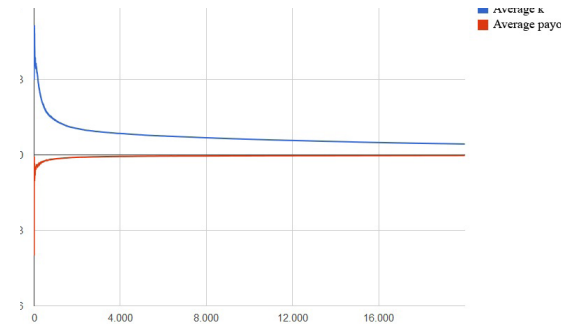


Figure 5: Simulation of the winner's curse situation over 20,000 time periods indicating the average strategy played and the average pay-off earned. It shows convergence towards optimal play.

This figure indicates that convergence towards the optimum indeed takes place. However, this rate of convergence is extremely sensitive on its initial condition of k and v . The figure also indicates the average pay-off that the participant received, which, too, shows a convergence towards the optimal pay-off, namely zero. In this specific situation, strategies 0,1,...,9 have been played 12,685, 6,453, 641, 164, 8, 6, 12, 8, 19, and 4 times respectively. The corresponding propensities are: 598 for strategy 0, 119 for strategy 1, 1 for strategy 2, and 0.0001 for all other strategies. The figure suggests a log-linear decrease of the average strategy and a log-linear increase of average pay-off.

The same can be plotted for shorter time periods, e.g. $t=500$ and $t=50$ as shown in Figure 6 and Figure 7. In both cases, convergence takes

place. In the second picture, the bidder plays an average strategy of 4 after 50 rounds. In this specific situation, the strategies 0,1,...,9 have been played 11, 4, 0, 4, 2, 11, 4, 10, 3 and 1 times respectively. The corresponding propensities are 10, 12, 8, 3, 2, 6, 12, 4, and 0.0001. Repeatedly plotting the same picture indicates an average strategy of 4 ± 1 over 50 time periods.

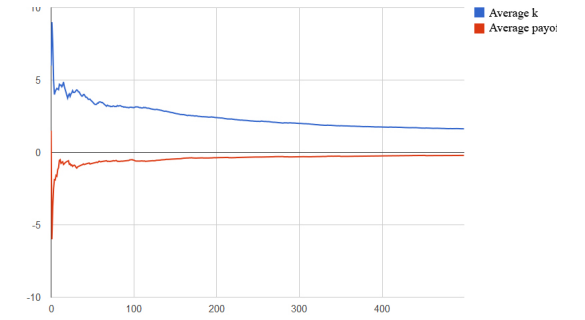


Figure 6: Simulation of the winner's curse situation over 500 time periods indicating the average strategy played and the average pay-off earned. It shows a convergence towards optimal play.

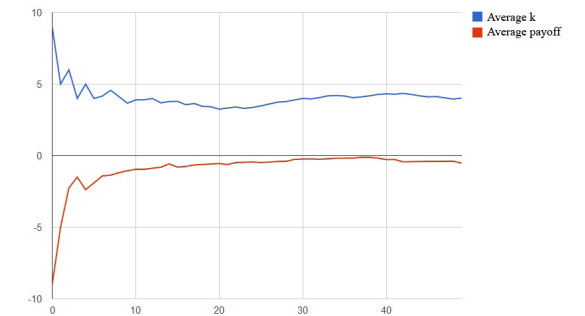


Figure 7: Simulation of the winner's curse situation over 50 time periods indicating the average strategy played and the average pay-off earned. It shows a convergence towards optimal play.

3.3.2. THE LOSER'S CURSE SITUATION

As for the winner's curse, Figure 8 gives the first intuition on what is happening in the loser's curse situation. The amount of Euros is indicated on the vertical axis and time periods are indicated on the horizontal axis. If the value exceeds the bid ($b = x + k$) and the bid is thus not accepted, the pay-off yields zero as in $t=1$ and $t=2$. From the picture it is clear that a higher bid results in a higher pay-off, as can be seen when comparing $t=0$, $t=3$ and $t=4$. In contrast to the winner's curse situation, the profits are never negative.

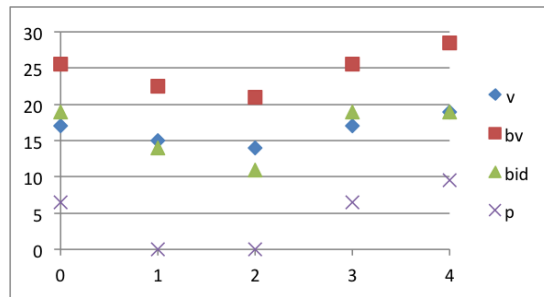


Figure 8: Simulation of the loser's curse situation over 5 time periods indicating the value (v), the buyer value (bv), the bid (b) and the pay-off (p) in Euros.

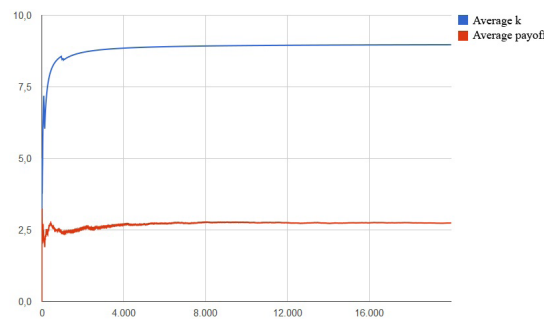


Figure 9: Simulation of the loser's curse situation over 20,000 time periods indicating the average strategy played and the average pay-off earned. It shows a convergence towards optimal play.

Learning can be determined by plotting the average strategy over time, as is done in Figure 9. This figure indicates that convergence towards the optimum indeed occurs. However, this rate of convergence has an extremely sensitive dependence on the initial conditions of k and v and furthermore, convergence only takes place 70% of the time. In the other 30% of the time no convergence, incomplete, or wrong convergence takes place. This will be discussed in the next section. If convergence to optimal play takes place, the rate of convergence seems higher than in the winner's curse situation. The figure also indicates the average pay-off that the participant has made, which also shows a convergence towards the optimal pay-off. The figure suggests a log-linear increase in the average strategy. In this specific situation, strategies 0,1,..9 have been played 6, 3, 21, 8, 25, 17, 24, 31, 3, and 19,862 times, respectively. The corresponding propensities are: 928 for strategy 9 and 0.0001 for all other strategies.

The same can be plotted for smaller time periods, e.g. $t = 500$ and $t = 50$ as shown in Figure 10 and Figure 11. In both cases, convergence occurs. In the second picture, the bidder plays an average strategy of 4 after 50 rounds. In this specific situation, strategies have been played 2, 1, 1, 1, 4, 14, 22, 1, 1, and 3 times, respectively. The corresponding propensities are 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 1, 15, 0.0001, 0.0001, and 0.04. From the propensities it indeed seems that learning takes place faster in the loser's curse situation.

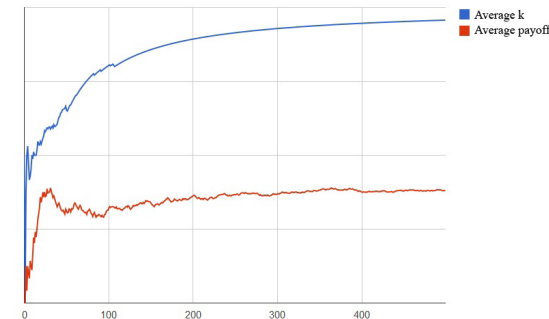


Figure 10: Simulation of the loser's curse situation over 500 time periods indicating the average strategy played and the average pay-off earned. It shows a convergence towards optimal play.

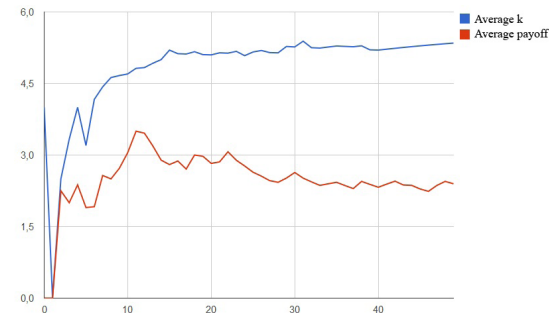


Figure 11: Simulation of the loser's curse situation over 50 time periods indicating the average strategy played and the average pay-off earned. It shows a convergence towards optimal play.

3.3.3. 'WRONG' LEARNING

As mentioned before, the convergence rate towards optimal play is extremely sensitive on the initial values of k and v . In the loser's curse scenario, however, this results in a negative or negligible convergence rate in 30% of the time. Often, 'wrong' learning or incomplete learning takes place, meaning that it converges upwards, but to a lower strategy than 9. Wrong learning can take place because the convergence is

mostly dependent on the results obtained in the first rounds. If strategies that are not optimal are positively reinforced in the beginning, the convergence is likely not to end up near to the optimal bidding strategy. Figure 12 shows two examples of wrong learning in the loser's curse situation. In the top panel, learning converges towards 6, indicating incomplete learning. In the bottom panel wrong learning takes place, as the average k converges downwards.

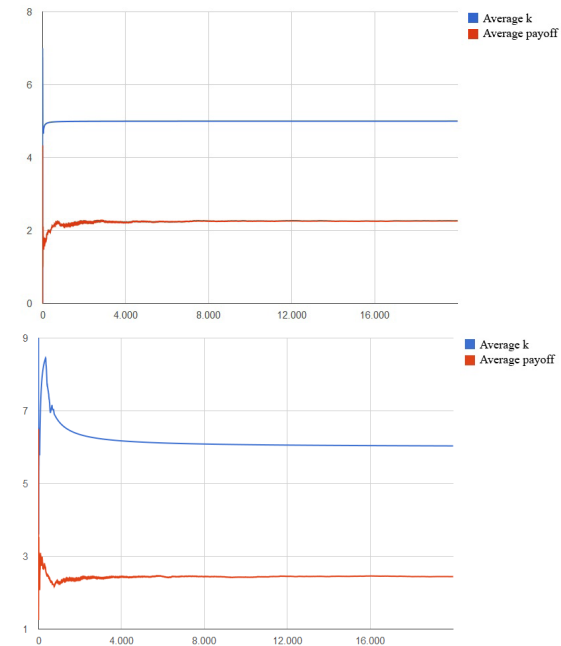


Figure 12: Simulation of the loser's curse situation over 20,000 time periods indicating the average strategy played and the average pay-off earned. In the top panel incomplete learning takes place, in the bottom panel wrong learning takes place.

Figure 13 shows the rare case of no convergence (top panel) and a situation in which the participant corrects for wrong learning after 4,000 rounds (bottom panel).

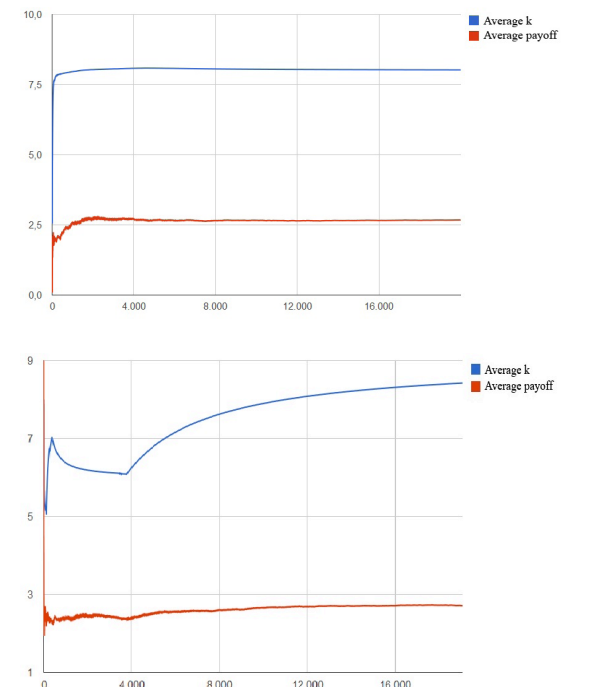


Figure 13: Simulation of the loser's curse situation over 20,000 time periods indicating the average strategy played and the average pay-off earned. In the top panel no convergence takes place, in the bottom panel the player corrects for wrong learning.

3.3.4. COMPARISON TO FELTOVICH'S RESULTS

In most cases, convergence seems to occur faster in my model than in Feltovich's model, as can be seen when comparing Figure 14 and Figure 15. The first figure shows Feltovich's convergence in winner's and loser's curse situations. The second figure illustrates my results. Due to the sensitive dependence on initial conditions and the values of the random draws of throughout the first few rounds, this picture cannot be generalized to all cases and only holds for the case with these particular initial conditions. Interestingly, Feltovich's winner's curse line is decreasing and concave down, whilst the winner's curse line in the second picture in this particular case is decreasing and concave up. It is unclear, however, whether Feltovich's results are based on a single player or multiple players participating in the one-player decision game. Since the intercept lies at $(0, 4.5)$, it is safe to assume that the results originate from multiple players, as playing $k = 4.5$ is not an option for

a single player in round $t = 0$. Section 4.3 discusses the graphical differences between both models more thoroughly.

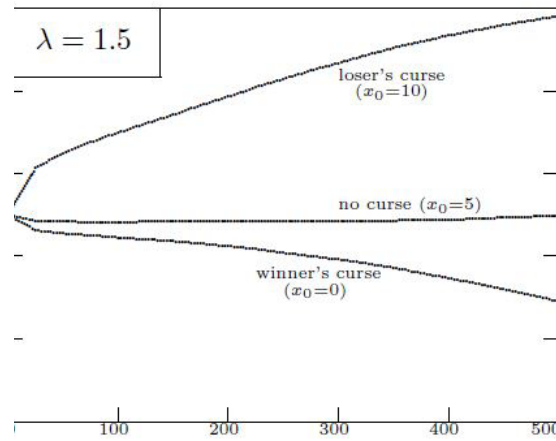


Figure 14: Results of simulations with fixed and varying x . The horizontal axis shows the time period and the vertical axis shows the average bid.

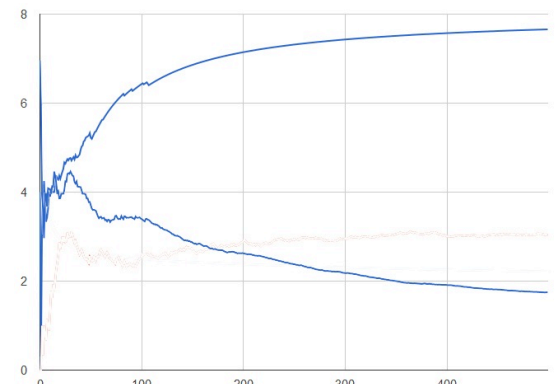


Figure 15: Feltovich's (2006) results of simulations with fixed and a varying x . The horizontal axis shows the time period and the vertical axis shows the excess of average bids.

3.4. CONCLUSION

In this chapter I have studied learning by plotting average strategies over time for individual players. Both in the loser's curse and winner's curse situation, convergence towards the optimum takes place for the majority of players. However, this rate of convergence has an extremely sensitive dependence on the first strategy chosen and the randomly drawn value of the object, and furthermore, convergence in the loser's

curse situation only takes place 70% of the time. In the other 30% of the time no convergence, incomplete, or wrong convergence takes place. If convergence to optimal play takes place in the loser's curse situation, the rate of convergence seems higher than in the winner's curse situation. The average convergence rate (averaged over many individuals) will be determined in Chapter 4.

Moreover, when comparing Feltovich's results to the results obtained by my simplified version of his model, convergence seems to take place faster in my model and the convergence of the winner's curse line has a different form, namely decreasing and concave up instead of decreasing and concave down. These results cannot be generalized, however, due to the sensitive dependence on initial conditions of especially the loser's curse scenario. Simulations of multiple players participating in the same one-player decision game must be examined in order to draw valid conclusions.

4. SIMULATIONS OF MULTIPLE PLAYERS

4.1. INTRODUCTION

In the previous chapter, we observed that the convergence rates in both winner's and loser's curse situations have an extremely sensitive dependence on the initial condition of k and v , and furthermore, that convergence in the loser's curse situation only takes place 70% of the time. In order to overcome this problem, this chapter runs the same simulation for 1,000 participants, instead of several individual players, and averages the results obtained by all participants. Section 4.2 reviews the results of these simulations and compares these results to the original results by Feltovich in Section 4.3. Section 4.4 determines the difference in convergence rates between the winner's and loser's curse situations and test whether this difference is statistically significant. Finally, Section 4.5 concludes.

4.2. RESULTS

4.2.1. THE WINNER'S CURSE SITUATION

Learning can be determined by plotting the average strategy over time and averaging the results of all 1,000 participants, as is done in Figure 16.

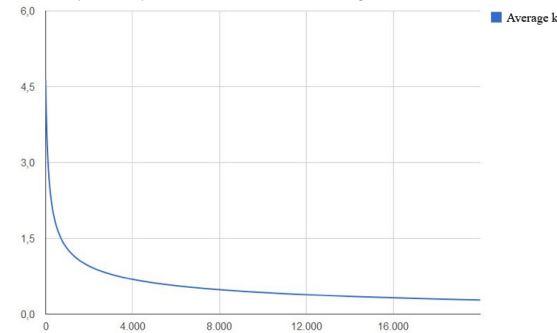


Figure 16: Simulation of the average winner's curse situation (1,000 participants) over 20,000 time periods indicating the average strategy played. It shows a convergence towards optimal play.

This figure indicates that convergence towards the optimum indeed takes place. The average player plays strategies 0,1,..9 in 16,512, 2,896, 409, 67, 49, 31, 32, 1, 1, and 2 of the rounds respectively. The corresponding propensities are: 613, 33, 0.1, 0.4, and 0.0001 for all other strategies. The figure suggests a log-linear decrease of average strategies.

The same can be plotted for smaller time periods, e.g. $t = 500$ and $t = 50$ as shown in Figure 17 and Figure 18. In both cases, convergence takes place. In the second figure, the average bidder plays an average strategy of 4 after 50 rounds. One average, the strategies 0,1,..,9 have been played 13, 8, 6, 5, 3, 1, 2, 3, 3, and 6 times, respectively. The corresponding propensities are 16, 13, 8, 4, 7, 2, 0.0001, 3, 0.5 and 5. The average convergence up to $t = 50$ indicates a linear downward sloping function. This will be investigated further in Section 4.4.1.

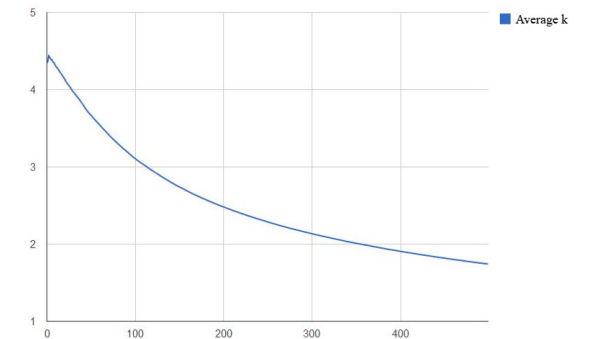


Figure 17: Simulation of the average winner's curse situation (1,000 participants) over 500 time periods indicating the average strategy played. It shows a convergence towards optimal play.

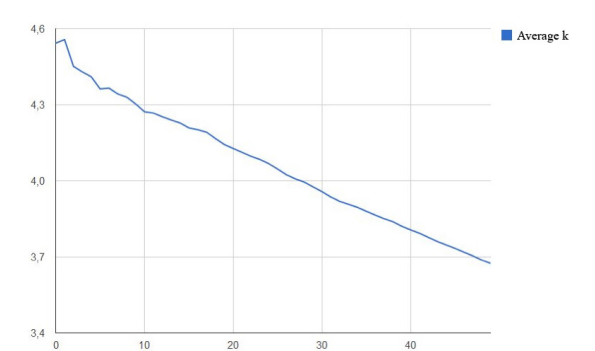


Figure 18: Simulation of the average winner's curse situation (1,000 participants) over 50 time periods indicating the average strategy played. It shows a linear convergence towards optimal play.

4.4.2. THE LOSER'S CURSE SITUATION

Learning can be determined by plotting the average strategy over time and averaging the results of all 1,000 participants, as is done in Figure 19. This figure indicates that convergence towards the optimum indeed takes place. The average player plays strategy 0,1,..9 in 2, 1, 1, 4, 47, 6,038, 1, 11, 80, 13,815 of the

cases, respectively. The corresponding propensities are: 2875 for strategy 9 and 0.0001 for all other strategies. The figure suggests a log-linear increase of average strategies.

The same can be plotted for shorter time periods, e.g. $t = 500$ and $t = 50$ as shown in Figure 20 and Figure 21. In both cases, convergence takes place. In the second picture, the average bidder plays an average strategy of 5.6 after 50 rounds. On average, the strategies 0,1,...,9 have been played 1, 1, 5, 9, 7, 2, 14, 3, 1, and 7 times, respectively. The corresponding propensities are 0.0001 for strategies 0, 1, 2, 6, 8, and 9 and 9, 5, 0.4, 3 for strategies 3, 4, 5, and 7. The average convergence up to $t = 50$ indicates a linear upward sloping function. This will be investigated further in Section 4.4.2.

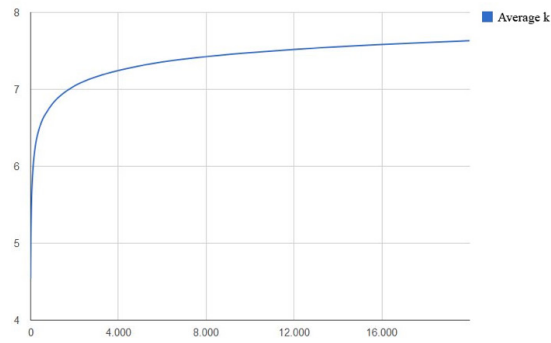


Figure 19: Simulation of the average loser's curse situation (1,000 participants) over 20,000 time periods indicating the average strategy played. It shows a convergence towards optimal play.

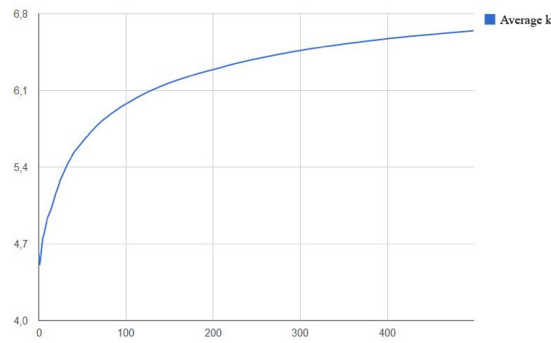


Figure 20: Simulation of the average loser's curse situation (1,000 participants) over 500 time periods indicating the average strategy played. It shows a convergence towards optimal play.

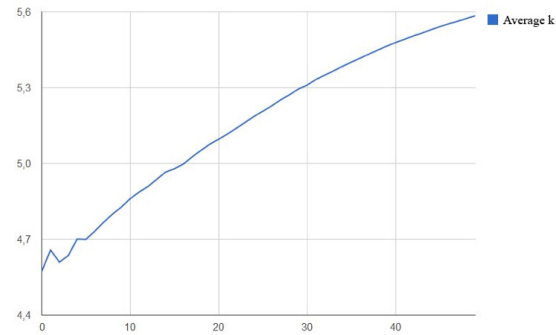


Figure 21: Simulation of the average loser's curse situation (1,000 participants) over 50 time periods indicating the average strategy played. It shows a linear convergence towards optimal play.

4.3. COMPARISON TO FELTOVICH'S RESULTS

Feltovich's graphical results, shown in Figure 22, suggest a higher convergence rate in the loser's curse situation than in the winner's curse situation.

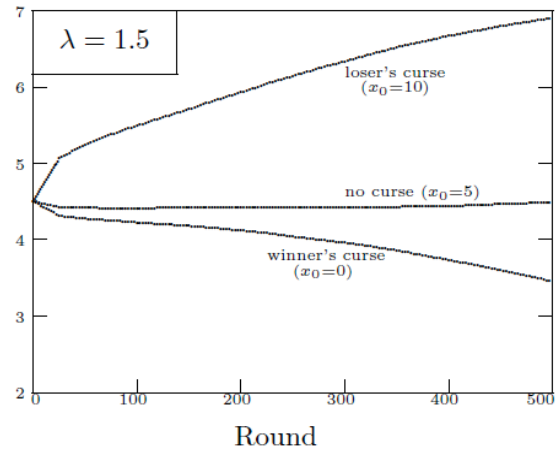


Figure 22: Feltovich's (2006) results of simulations with fixed and a varying λ . The horizontal axis shows the time period and the vertical axis shows the excess of average bids.

This statement is solely based on visual inspection and has neither been mathematically proven nor has Feltovich tested whether the difference in convergence between both curses is significantly different. My results, shown in Figure 23, seem to suggest the direct opposite: the winner's curse converges faster than the loser's curse.

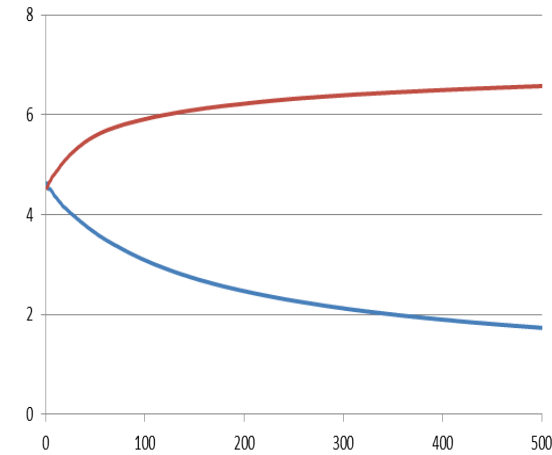


Figure 23: Average of 1,000 simulations with fixed and varying x in loser's (red) and winner's (blue) curse situations. The horizontal axis shows the time period and the vertical axis shows the average bid an average player submits.

In comparison, an average participant in the loser's curse situation ends up playing an average strategy of approximately 6.75 after 500 rounds in both models. In the winner's curse situation, an average participant ends up playing an average strategy of 3.5 in Feltovich's model and 1.75 in my simulation.

In order to investigate which change in parameter values or reinforcement accounts for the greatest difference between Feltovich's model and the model discussed in this thesis, the next subsections will reintroduce parameters to the model and isolate their individual effects on the learning outcomes.

4.3.1. THE INFLUENCE OF GRADUAL FORGETTING AND EXPERIMENTATION

When reintroducing the parameter experimentation between strategies that the averages player engages in ($\epsilon = 0.1$) to the model, this has little effect on the convergence rates of the loser's and winner's curse graph. Figure 24 (top) shows that the experimentation parameter has a negligible negative effect on the winner's curse, with an average deviation of 0.020 from the graph without this parameter. This parameter has about 5% more (positive) effect on the loser's curse, with an average deviation of 0.117.

The influence of reintroducing the gradual forgetting parameter ($\delta = 0.001$) to the model can be seen in Figure 24 (middle panel). The opposite effect is present here: the gradual forgetting parameter has about 10% more (positive) influence on the

winner's curse than on the loser's curse (negative effect), with a deviation of -0.078 from the winner's curse and -0.009 loser's curse graph. In total, the gradual forgetting parameter has little effect on the convergence rate in both situations.

Similarly, reintroducing both the experimentation and gradual forgetting parameters to the model has little positive effect on the convergence rates, with a deviation of -0.089 from the winner's curse and 0.092 from the loser's curse graph. Thus, the gradual forgetting and experimentation parameters are not an important cause of the difference between my model and Feltovich's model. These results are summarised in Table 4.

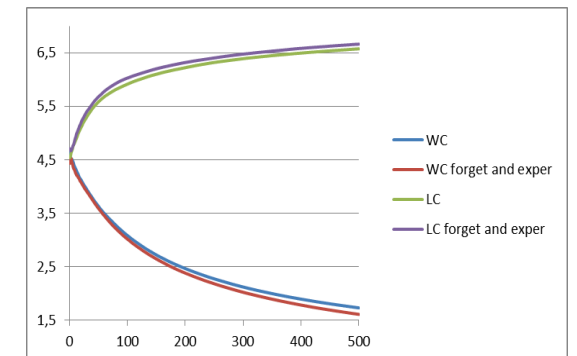
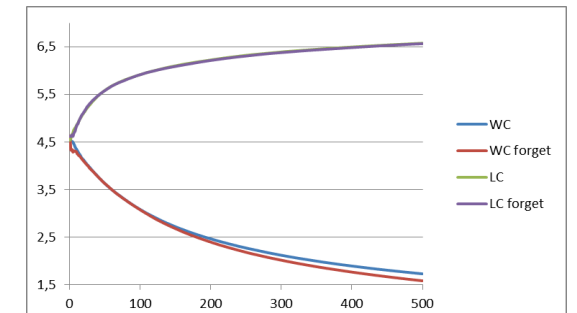
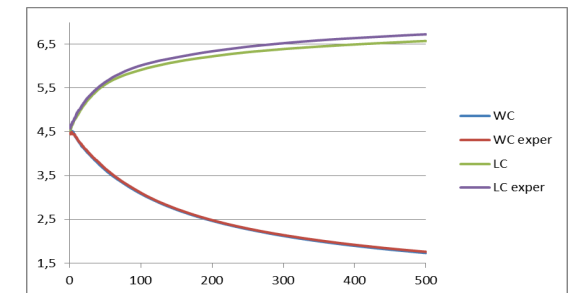


Figure 24: The effects of adding the experimentation parameter (top), the gradual forgetting parameter (middle) and the experimentation and gradual forgetting parameters (bottom) to the model.

Average differences	With the gradual forgetting parameter	With the experimentation parameter	With gradual forgetting and experimentation parameters
Winner's Curse	-0.078	+0.020	+0.089
Loser's Curse	-0.009	+0.117	+0.092

Table 4: The average effect of adding the gradual forgetting parameter, the experimentation parameter and the gradual forgetting and experimentation parameters to the model in winner's and loser's curse situations.

4.3.2. THE INFLUENCE OF REINFORCING NEIGHBOURING STRATEGIES

Reintroducing to the model the reinforcement of the propensities associated with the neighbouring strategies of the strategy that was played yields that similar strategies are now also slightly reinforced, according to the following formulas:

$$q^{t+1}(k) = (1-\delta)q^t(k) + (1-\varepsilon) \left[\prod^t(k,v) - \rho^t \right]$$

$$q^{t+1}(j) = (1-\delta)q^t(j) + \frac{\varepsilon}{M} \left[\prod^t_n(k,v) - \rho^t_n \right] \text{ for } |j-k|=1$$

$$q^{t+1}(j) = (1-\delta)q^t(j) \text{ for } |j-k|>1$$

Recall that M represents the number of neighbouring strategies. If $k=0$ or $k=9$ than $M=1$, $M=2$ in all other cases. I will first examine the differences without including the experimentation and gradual forgetting parameters parameters ($\varepsilon=0$ and $\delta=0$); later I will reintroduce both parameters into the model.

Reintroducing the reinforcement of neighbouring strategies to the model has little effect on the convergence rates of the loser's and winner's curse graph. Figure 25 (top panel) shows that this has a negligible negative effect on the winner's curse, with an average deviation of 0.001 from the graph without this reinforcement. The reinforcement of neighbouring strategies has a greater (positive) effect on the loser's curse, with an average deviation of 0.089.

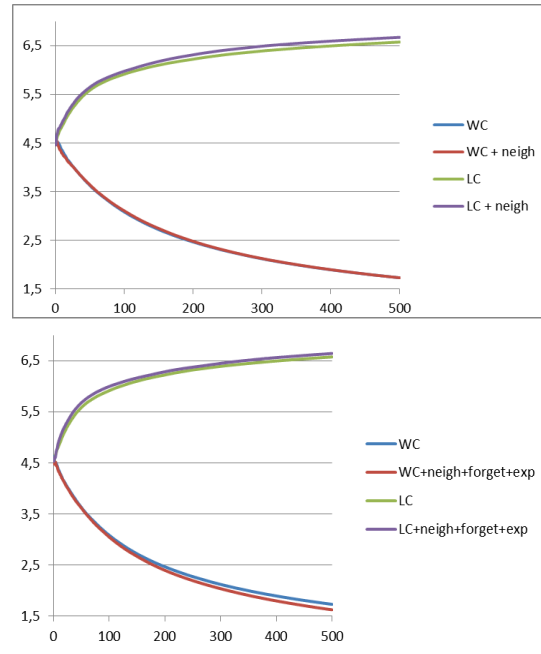


Figure 25: The effects of reinforcing neighbouring strategies (top) and reinforcing neighbouring strategies including the experimentation and gradual forgetting parameters (bottom) in the model.

Reintroducing both the reinforcement of neighbouring strategies and the experimentation and gradual forgetting parameters to the model has a similar absolute effect on both the convergence rates of the winner's curse, with an average deviation of -0.069, and the loser's curse, with an average deviation of 0.068. These results are summarised in Table 5.

Average differences	With $\varepsilon=0.1$ and $\delta=0.001$ (from Section 4.3.1)	With reinforcement of neighbouring strategies with $\varepsilon=0$ and $\delta=0$	With reinforcement of neighbouring strategies with $\varepsilon=0.1$ and $\delta=0.001$
Winner's Curse	-0.089	+0.001	-0.069
Loser's Curse	+0.092	+0.086	+0.068

Table 5: The average effect of reinforcing neighbouring strategies and reinforcing neighbouring strategies whilst including the experimentation and gradual forgetting parameters in the model in winner's and loser's curse situations.

4.3.3. THE INFLUENCE OF USING A DIFFERENT REFERENCE VALUE

So far I have considered the influence of reintroducing the gradual forgetting parameter, the experimentation parameter and the reinforcement of neighbouring strategies (or a combination of the three) into the model. As shown above, the effects were minor, if

not negligible. This indicates that the main difference between Feltovich's and my results must be due to one of the other simplifying assumptions made to the model, as discussed in Section 3.2.3. I will now examine the effect of using the average payoff as reference value. Thus, I will reintroduce the reference value used by Feltovich, given by:

$$\rho^{t+1} = \begin{cases} (1-w^+)(\rho^t) + (w^+) \prod^t(k,v), & \prod^t(k,v) \geq \rho^t \\ (1-w^-)(\rho^t) + (w^-) \prod^t(k,v), & \prod^t(k,v) < \rho^t \end{cases}$$

Feltovich sets the initial reference point equal to the minimum possible pay-off, $(\lambda-1)x-9$, yielding in the winner's curse and -4 in the loser's curse situation. As discussed before, I find this counter-intuitive, since few people will enter an auction if they expect to make negative profits in the first place. I will therefore set the initial reference point equal to zero. Recall that w^+ represents outcomes that are better than expected and w^- represents outcomes that are worse than expected. I will first consider the case in which $w^+ = w^- = 0$, later I will reintroduce both parameters into the model ($w^+ = 0.02$ and $w^- = 0.01$). The results are summarised in Table 6.

Reintroducing the original reference value to the model has a significant effect on both convergence rates of the loser's and winner's curse graph. Figure 26 (top panel) shows that this has a large positive effect on the winner's curse, with an average deviation of -0.558 from the graph without this alteration. The reinforcement of neighbouring strategies has a smaller (negative) effect on the loser's curse, with an average deviation of -0.275.

Reintroducing the parameters $w^+ = 0.02$ and $w^- = 0.01$ into the model slightly reduces the effect of the original reference value, shown in Figure 26 (bottom panel). The average deviation in the winner's curse situation is -0.526 and in the loser's curse situation is -0.268.

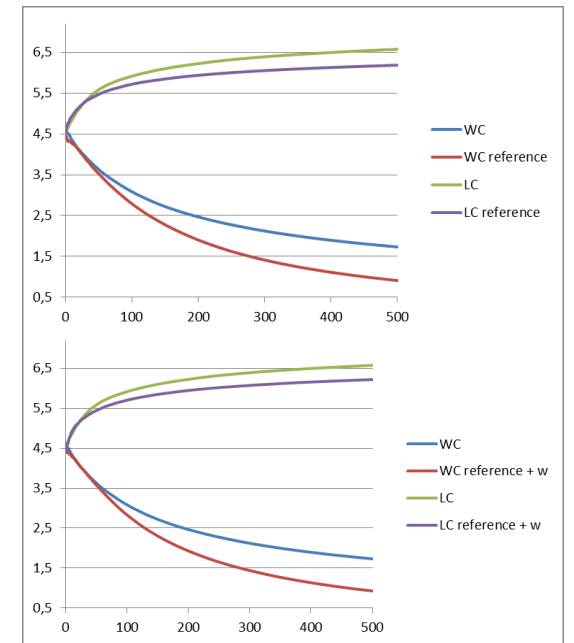


Figure 26: The effects of using a different reference value (top) and using a different reference value including $w^+ = 0.02$ and $w^- = 0.01$ (bottom) in the model.

Interestingly, when comparing the graph in Figure 27 to the graph in Feltovich's paper, the differences have become even more striking. Whilst Feltovich's graph indicates a faster divergence in the loser's curse situation, Figure 27 shows precisely the opposite, meaning that one of the two other changes made to the model should account for this difference. This can be seen in Figure 28.

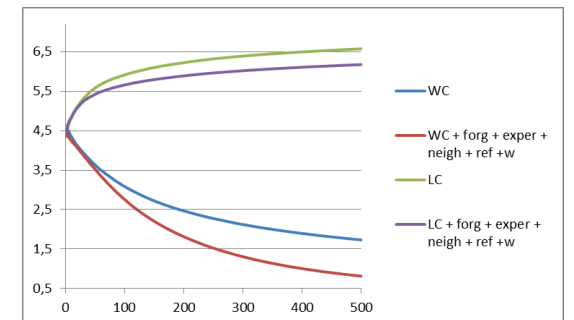


Figure 27: The effect of reintroducing all effects discussed so far into the model.

Average differences	Effect of using a different reference value	Effect of using a different reference value with $w^+ = 0.02$ and $w^- = 0.1$	Effect of using a different reference value with $w^+ = 0.02$ and $w^- = 0.1$ and all other effects discussed so far
Winner's Curse	-0.558	-0.526	-0.635
Loser's Curse	+0.275	-0.268	-315

Table 6: The average effect of using a different reference value in combination with other effects.

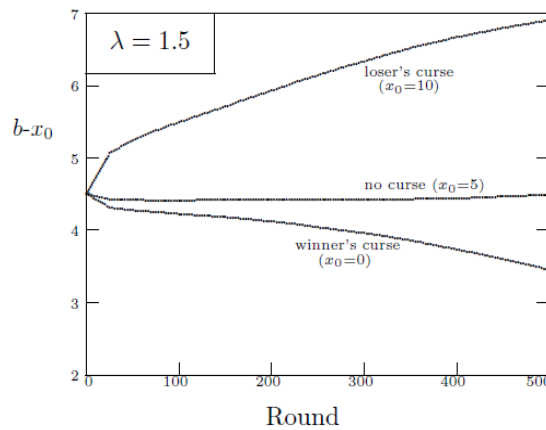


Figure 28: In the top panel, Feltovich's (2006) results of simulations with fixed λ and a varying x . In the bottom panel, the results of reintroducing all effects discussed so far into the model.

4.3.4. OTHER INFLUENCES

The effects of the alterations to the Feltovich's model that still need to be considered include the effect of setting a different initial propensity and the effect of introducing a minimum propensity value to the model.

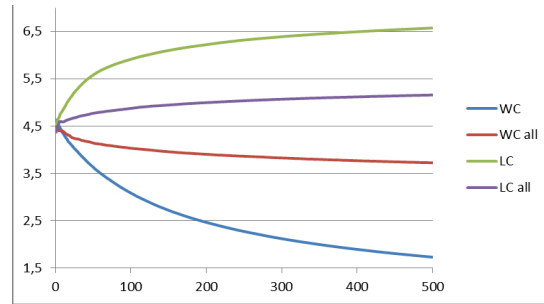


Figure 29: Effect of reintroducing everything so far including the initial reference value into the model.

Recall that Feltovich uses $(\lambda - 1)x - 9$, the minimal possible pay-off, as initial reference value, yielding -9 in the $x = 0$ case and -4 in the $x = 10$ case. Reintroducing this into the model results in graph shown in Figure 29. This minor change turns out to have a great effect on the model's outcome, as it decrease the convergence rates of the winner's and loser's curse tremendously. These results are summarised in Table 7.

Average differences	Effect of reintroducing everything so far including the initial reference value
Winner's Curse	-1.405
Loser's Curse	-1.179

Table 7: The average effect of reintroducing everything so far including the initial reference value.

In order to ensure that the propensities do not become negative, as this would yield negative probabilities, I have set all initial propensities to 5, and propensities that are reinforced below 0.001 are automatically set back to 0.001. Feltovich uses $Q = (\lambda - 1)x$ as initial value for the summation of all propensities, which is roughly the average magnitude of payoffs. As propensities are equal in the first round, this would yield 0 and 0.5 in $x = 10$ for all ten propensity values. In this case, a negative reinforcement at the beginning of the simulation would result in negative probabilities and unrepresentative results. Unfortunately, it is impossible to reintroduce this into the model, as the simulation cannot account for negative propensities. However, I expect that this will alter the outcomes significantly, explaining why Figure 29 and Feltovich's results still differ to a large degree.

4.4. CONVERGENCE RATES TOWARDS OPTIMAL BIDS

In order to determine the rates of convergence towards optimal play in winner's and loser's curse situations and determine which of both converges faster, the best fitted (linear) function is used as an approximation to the data. The convergence rates for $t = 50$ and $t = 20,000$ are investigated, as the first represents the maximum number of trials in which people in a laboratory usually participate and the later represents long-term behaviour.

4.4.1. THE WINNER'S CURSE SITUATION

The learning rate of the first 50 trials of the winner's curse can be best fitted by a linear ordinary least squares (OLS) regression of the form $\bar{b} = -0.02t + 4.54$ meaning that the average player's average strategy played (\bar{b}) moves about 0.02 in the direction of optimal play every round (t). This can be seen in Figure 30.

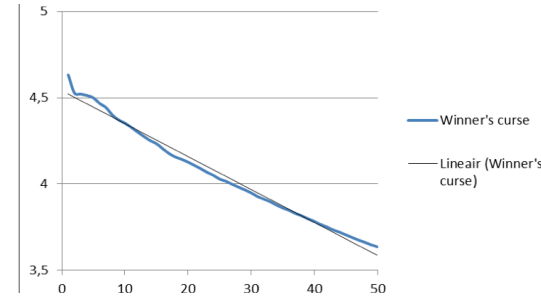


Figure 30: Simulation of the average winner's curse situation (1,000 participants) over 500 time periods indicating the average strategy played and its best linear fit.

The average strategy curve of the average player participating in 20,000 rounds of the game in the winner's curse situation is best fitted using OLS by the log-linear function $\bar{b} = -0.40\ln(t) + 4.15$. This can be seen in Figure 31.

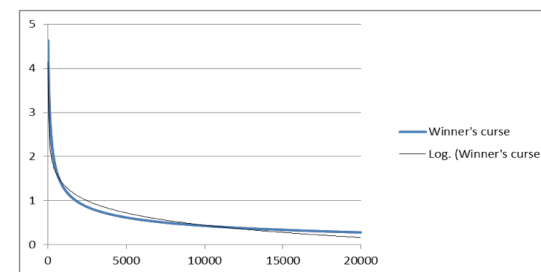


Figure 31: Simulation of the average winner's curse situation (1,000 participants) over 20,000 time periods indicating the average strategy played and its best log-linear fit.

4.4.2. THE LOSER'S CURSE SITUATION

The learning rate of the first 50 trials of the winner's curse can be best fitted using OLS by a linear function of the form $\bar{b} = 0.02t + 4.65$, meaning that the average player's average strategy played moves about 0.02 in the direction of optimal play every round. This can be seen in Figure 32.

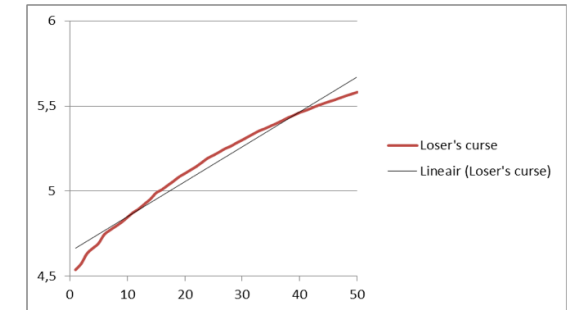


Figure 32: Simulation of the average loser's curse situation (1,000 participants) over 500 time periods indicating the average strategy played and its best linear fit.

The average strategy curve of the average player participating in 20,000 rounds of the game in the loser's curse situation, is best fitted using OLS by the log-linear function $\bar{b} = 0.29\ln(t) + 4.78$. This can be seen in Figure 33.

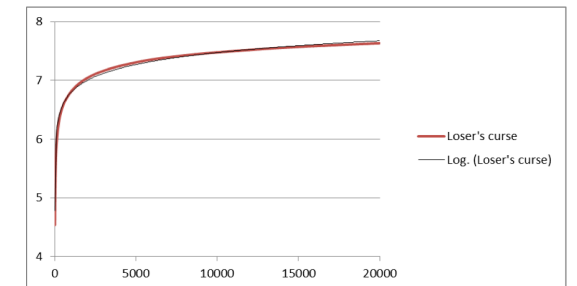


Figure 33: Simulation of the average loser's curse situation (1,000 participants) over 20,000 time periods indicating the average strategy played and its best linear fit.

4.4.3. COMPARISON BETWEEN LOSER'S AND WINNER'S CURSE

The results of the fittings done in the previous two sections are summarised in Table 8. In order to determine which of the curses converges faster towards optimal play, these fittings must be compared to each other.

For the first 50 rounds, both curses can be estimated by a linear fitting. Comparing both lines

indicates that the loser's curse converges faster than the winner's curse, providing evidence in support of my hypothesis. Interestingly, when comparing the log-linear fittings for the 20,000 rounds situation, the opposite conclusion can be drawn: the winner's curse converges faster towards optimal play in the long-term than the loser's curse does.

The OLS method, however, is not suitable to test whether these results are significant, as one of the assumptions underlying OLS estimation is that the errors are uncorrelated. In time series data, and especially in such a feedback system, this assumption is easily violated. A variation to the difference-in-difference (DID) is a more appropriate method.

Fitted function	50 rounds	20,000 rounds
Winner's curse	$\bar{b} = -0.02t + 4.54$	$\bar{b} = -0.40\ln(t) + 4.15$
Loser's curse	$\bar{b} = 0.02t + 4.65$	$\bar{b} = 0.29\ln(t) + 4.79$

Table 8: Fitted function of the winner's and loser's curse curves for $t=50$ and $t=20,000$.

For all the time periods in the sets $\{1, \dots, 50\}$, $\{451, \dots, 500\}$, $\{1001, \dots, 1050\}$ and $\{19951, \dots, 20000\}$, the strategies played by 1,000 participants in the winner's (w) and loser's (l) curse treatment are averaged and are given by b_{it} with $i = \{w, l\}$ and $t \in \{1, \dots, 50\}, \{451, \dots, 500\}, \{1001, \dots, 1050\}$ and $\{19951, \dots, 20000\}$. $Y_{it} = |b_{it}^{opt} - b_{it}|$ represents the absolute difference between the optimal bid and the average strategy played by 1,000 participants per treatment for all time periods in the interval. The relationship between the type of curse and this absolute difference is given by:

$$Y_{it} = \alpha_T + \beta_T D_{it} + \epsilon_{it}$$

With D_{it} is a dummy variable which indicates whether it is a winner's $D_{it} = 0$ or a loser's $D_{it} = 1$ curse situation. α_T roughly equals Y_{it} when $D_{it} = 0$, meaning that α_T is the value of the absolute difference between the optimal bid and the average strategy played by 1,000 participants participating in the winner's curse treatment at a certain time period. β_T represents the difference between Y_{it} and α_T , thus $Y_{it} - Y_{wT}$. For the average time period in $\{1, \dots, 50\}$, OLS estimates $Y_{it} = 3.68 - 0.33D_{it}$

, and $Y_{it} = 1.04 + 1.04D_{it}$, $Y_{it} = 0.73 + 1.03D_{it}$ and $Y_{it} = 0.09 + 1.09D_{it}$ for the second, third and fourth sets respectively. This yields an insignificant difference in convergence rates between the winner's and the loser's curse in the first time set and significant differences in the other time sets at a 95%

significance level. Thus, the winner's curse converges significantly faster than the loser's curse.

4.5. CONCLUSION

In order to account for the extreme sensitive dependence of the simulations of one player participating in a one-player decision game in winner's and loser's curse context, this chapter runs a simulation for 1,000 participants and averages the results obtained by all participants.

The convergence rates for $t = 50$ and $t = 20,000$ are investigated, as the first represents the maximum number of trials in which subjects in a laboratory usually participate and the latter represents the long-term behaviour. The first 50 trials can be fitted by a linear function in both winner's and loser's curse situation. In accordance with my hypothesis, the loser's curse converges faster than the winner's curse. However, using a difference-in-difference approach it can be concluded that this difference is not statistically significant. In the long-term, for $t = 20,000$, both winner's and loser's curse can be fitted with a log-linear function, indicating that the winner's curse converges faster than the loser's curse. This results contradicts my and Feltovich's hypothesis. The difference in convergence rates is statistically significant, which was also determined using a difference-in-difference approach.

In order to investigate which change in parameter values or reinforcement accounts for the greatest difference between Feltovich's model and my model, I reintroduced these parameters to the model and isolated their individual effects on the learning outcomes. Reintroducing the experimentation parameter, the gradual forgetting parameter and the reinforcement of neighbouring strategies in the model, has a negligible effect on the outcomes. Reintroducing the reference value and its initial condition as used by Feltovich, however, has a tremendous effect on the outcomes in an unexpected direction.

5. CONCLUSION

This final chapter will first summarize the main results and conclusions presented throughout this thesis (Section 5.1). Section 5.2 discusses the implications of these results for businesses and briefly hints at some suggestions for further research that would create a more realistic business context. These and more suggestions for further research will be discussed thoroughly in Section 5.3 and include both suggestions for additions to the model and for experimental testing in order to determine which effect dominates empirically.

5.1. SUMMARY

This thesis investigated whether the loser's curse is more persistent than the winner's curse in a one-player decision model based on Bazerman and Samuelson's (1985) buy-a-firm problem using a simplified version of Feltovich's (2006) reinforcement learning model. It hypothesized the following: 1) slow or no convergence to optimal play occurs in both the winner's and loser's curse situations and 2) convergence to optimal play in the loser's curse situation takes place significantly faster than in the winner's curse situation.

Experimental and field evidence both indicate that the winner's curse and loser's curse exist and persist, and that people thus often fail to take the asymmetric information that is built into the problem into consideration.

In my simulations of the buy-a-firm problem in winner's and loser's curse context, I observed a convergence towards optimal play for individual players, however, this rate of convergence has an extremely sensitive dependence on the first strategy chosen and the randomly drawn value of the object, and furthermore, convergence in the loser's curse situation only takes place 70% of the time. In the other 30% of the time no convergence, incomplete, or wrong convergence takes place.

Averaging the simulations of 1,000 players participating in the buy-a-firm problem, I observed the following: The first 50 trials can be fitted by a linear function in both winner's and loser's curse situations. In accordance with my hypothesis, the loser's curse converges faster than the winner's curse. However, this difference is not significant when using a difference-in-difference method. In the long-term, for $t = 20,000$, both winner's and loser's curse

can be fitted with a log-linear function, indicating that the winner's curse converges faster than the loser's curse. These results contradict my and Feltovich's hypothesis. This difference in convergence rates is shown to be statistically significant, when using a difference-in-difference approach.

To sum up, I only found partial support for the hypotheses presented in the introduction of this thesis. I found support for hypothesis 1, slow or no convergence to optimal play occurs in both the winner's and loser's curse situations, as in both situations players need over 500 rounds to learn to play optimally. The situation in which 20,000 trials are used also indicates that convergence is extremely slow. The support for hypothesis 2, convergence to optimal play in the loser's curse situation takes place significantly faster than in the winner's curse situation, is less strong. The first 50 rounds suggest that the hypothesis holds, the long-term behaviour suggests the opposite. It is left to further research to determine which effect dominates empirically.

5.2. DISCUSSION: IMPLICATIONS FOR BUSINESSES

All of us, including businesses, are susceptible to the winner's and loser's curse. We can easily overpay for something that we passionately want to acquire or underbid due to risk aversion or conservative tendencies. These curses do not occur with every decision that we make, but both curses occur often enough to form a great threat to businesses. Especially given the prevalence of auctions in business today, it is extremely important for decision-makers (i.e., bidders or managers in this case) to fully understand the nature of the winner's and loser's curse, and to be aware of their own irrational behaviour, as over- and underbidding can have disastrous consequences for the earnings and the costs of a business (Warren and Samuel, 2009). Note that the occurrence of these curses is not limited to auctions; it can happen in every situation where there is uncertainty about a value of an object. Today's managers are more and more faced with decisions involving many countries, due to increasing globalisation. Thus, the opportunities to encounter the winner's and loser's curse are expending every day (Anandalingam and Lucas, 2004).

According to economists, the main reason that subjects fail to recognise the profit-maximizing decision strategy (the rational choice), is because

they are unable to properly consider all available information (Savikhin et al., 2006). The forces that promote the curses include psychological and personal factors, coming from within an individual or organisation, and market factors, which are largely external for the decision-maker. The first includes, among others, buying and bidding psychology (getting caught up in the excitement of an auction), competition and winning (winning is everything), risk aversion (scared of losing money), and managerial optimism (everything will turn out fine). The second includes, among others, pressure to grow and unrealistic business models. Obviously, not all of these forces apply to every situation (Anandalingam and Lucas, 2004). The psychology underlying agent's irrational behaviour goes beyond the scope of this thesis, but is nevertheless important to keep in mind.

In the simulations developed in Chapters 3 and 4, we have seen that, if agents indeed learn via reinforcement, it will take them over 20,000 rounds to learn how to play the optimal (rational) bid. The model is based on a simplified framework, the buy-a-firm problem, in which subject do not interact with one another, but instead bid on a company with a randomly determined value unknown to the bidder. In reality, auctions often involve multiple players, meaning that bidders do not only need to strategically determine their bids, but also need to compete against one another to win the company. If competitors, i.e. other bidders in the same auction, were included in the model, agents would learn to overcome the winner's curse even slower, as they only receive feedback when they 'win' the auction. Obviously, the chance of winning the auction is significantly smaller when there are more competitors. In contrast, agents would learn to overcome the loser's curse more quickly when more competitors are added to the model, as the chance that they lose the auction increases. This might change the result that the loser's curse converges faster than the winner's curse in the first 50 rounds and vice versa after 20,000 rounds.

Most likely, the above-mentioned addition to the model would shift the results and provide stronger evidence for the second hypothesis proposed in this thesis, namely that people learn to overcome the loser's curse more quickly than they learn to overcome the winner's curse. This would suggest that a company's priority should be on learning to

overcome the winner's curse, as this curse is the most persistent and the most vicious of the two curses, and thus more detrimental to business.

As already discussed in Section 2.3, many studies have run experimental trials to address the winner's and loser's curse, mostly based on the buy-a-firm problem, in order to find a way for subjects to avoid both curses by attempting to improve their decision-making abilities. Idson et al. (2004), for example, trained their participants in solving various conditional probability problems before running their experiment. However, this training was only marginally effective in improving learning. Thus, simply giving your employees training in probability theory will not help to avoid the loser's or winner's curse. The influence of training on decision-making abilities also goes beyond the scope of this thesis, but might be relevant for further research.

In the book *Beware The Winner's Curse* (Anandalingam and Lucas, 2004), economic game theory and scenario planning are proposed as useful mechanisms to avoid the winner's curse. The first ensures that a company also analyses the decision through the eyes of its competitors, if applicable, and the seller of the object. The second provides companies with the tools to deal with uncertainty. Furthermore, groups are more likely to reach an optimal bidding strategy than individual bidders, as pointed out by the experiment conducted by Casari, Jackson, and Zhang (2009). A company might benefit by letting a group of people make the decisions instead of a single expert. All these proposals help to reduce the risks of falling prey to one of the curses. However, since the understanding of the mechanisms underlying the winner's and loser's curse is still in development and more research is yet to be done, being aware of the curses and their effects is the first step towards avoiding them.

5.3. SUGGESTIONS FOR FURTHER RESEARCH

I believe that more research needs to be done in this area. The implications of the model I used throughout this thesis need to be tested further mathematically and empirically, and should also be compared to other learning models. Mathematical additions to the model (Section 5.3.1) and experimental treatments (Section 5.3.2) can help to test and increase the ability of this model to describe individual behaviour in decision-making games. Given the prevalence of

auctions in business today, it is extremely important for decision-makers to fully understand the nature of the winner's curse and loser's curse, as over- and underbidding can have disastrous consequences on individual budgets and businesses. More experimental research might help shed light on the reasons for the persistence of irrational behaviour in this context.

5.3.1. ADDITIONS TO THE MODEL

As I have already pointed out in Section 5.2, the buy-a-firm game studied in this thesis is a highly simplified version of reality. This decision-making problem captures the essential features of a one-player decision game, but has a questionable connection to real-life markets in which businesses operate, as the model assumes that interactions solely take place between one bidder and one seller, in which the seller makes no strategic decisions and simply accepts a bid if it is equal to or exceeds the value of the object.

A more realistic model would include multiple bidders and a seller making strategic decisions. Probably, it would also include multiple objects to be sold, and perhaps even multiple sellers. As hypothesized above, allowing these additions to the model would not invalidate the results obtained in this thesis and would most likely provide stronger evidence for faster convergence in the loser's curse than in winner's curse situation. In that case, a larger proportion of bids would be unsuccessful, leading to zero pay-offs rather than negative pay-offs in the winner's curse situation (meaning less learning opportunities) and more learning opportunities in the loser's curse situation. The propensities in the winner's curse context would then change even more slowly over time. In more realistic markets it is, therefore, likely that hypotheses 1 and especially 2 are supported more strongly than is currently the case. Obviously, this still needs to be tested by adding the proposed additions to the simulation.

In the existing model, parameter values can be varied to determine the individual effect which every parameter has on the outcomes of the model. This is useful for optimizing the model's representativeness for empirical data and to test whether the assumptions made in the model are valid.

Different forms of learning models have become increasingly popular in the scholarly literature, all with their own strengths and weaknesses and applicability to different situations

(ranging from economic to biological applications). This thesis assumes reinforcement learning. It would be interesting to see the differences in outcomes when applying different learning models. A comparison to empirics can help to determine which learning method is most applicable in the winner's and loser's curse situation and can, thus, help shed light on the reasons for the persistence of irrational behaviour in this context.

5.3.2. EXPERIMENTAL TESTING

In order to determine whether the winner's or loser's curse dominates empirically and to test whether the reinforcement learning model used throughout this thesis is a good representation of human behaviour (inside and outside of the economic laboratory), a laboratory experiment should be designed that captures the essential features of this decision-making problem. I propose the following experimental design:

The experiment involves 3 trials per treatment group (6 trials in total). The first treatment group consists of 20 participants engaging in 50 rounds of the buy-a-firm problem in the winner's curse context per trial. The second treatment group consists of a similar number of participants and rounds in the loser's curse context per trial. All subjects participating are bidders. The feedback that subjects receive after every round is similar in both treatments.

The goal of the game is not to win the object, i.e. the auction, as many times as possible, but is instead to maximize profits. Several test questions should be included before the game starts to test whether the participants have understood the game, as failure to do so is detrimental to the outcomes. Furthermore, it must be understood that the seller does not act according to a strategic profit maximizing plan, and accepts every bid equal to or greater than the value of the object.

Since the winner's curse situation is designed in such a way that bidding optimally yields zero pay-off and the loser's curse situation is designed to receive a strictly positive pay-off, both treatments require a different pay-out system in order to make sure that participants have sufficient incentives to play the game seriously. The expected pay-out in both treatments should be equalized, meaning that participants in the winner's curse situation should

start with a high initial budget (as they will most likely lose money) and participants in the loser's curse situation should start with a low initial budget (as they are likely to win money). At the end of the 50 rounds, participants will receive a pay-out equal to their account in round 50. The minimum pay-out equals the standard show-up fee. One must keep in mind, however, that the differences in how participants experience losing money or winning money might affect their playing strategies. I realize that this pay-out mechanism is not optimal, but it ensures that players have incentives to learn and participate seriously.

Which forms of feedback are used in the experiment is dependent on the research question. The forms of feedback that should always be included in the experiment are the bid that the participant placed, the value that the object had and the profit that the participant received. Additional feedback, for example, might include a screen that displays the average pay-off. The model used in this thesis assumes a reference value equal to the average pay-off. As mentioned before, 'thinking in averages' might be problematic in the long-term, and therefore not completely representative to the real world. I assumed people to be capable of calculating the average over longer time periods. Providing the average pay-off as feedback makes it possible to test whether people who indeed have this information (or are capable of calculating it) will also use this information in their learning process. For a different research questions, however, it might be better not to include this information, e.g. when testing what kind of learning takes place.

The same experimental design can later be extended to account for the additions to the model proposed in Section 5.3.1 and to create a more realistic representation of the context in which businesses operate. Further research will greatly contribute to a full understanding of the nature of markets with asymmetric information and the mechanisms that underlie the winner's curse and loser's curse, and will help individuals and companies to become aware of their irrational behaviour that might have disastrous consequences.

REFERENCES

- Anandalingam, G. and Lucas, H. (2004). *Beware the Winner's Curse: Victories That Can Sink You and Your Company*. New York: Oxford University Press.
- Ball, S., Bazerman, M. and Carroll, J. (1991). An Evaluation of Learning in the Bilateral Winner's Curse. *Organizational Behavior and Human Decision Processes*, Vol. 48 No. 1, 1-21.
- Beggs, J. (2010, April 21). *The Winner's Curse, Olympics Edition... [sic]*. Retrieved December 18, 2011, from Economists Do it With Models: <http://www.economistsdoitwithmodels.com/2010/04/21/the-winners-curse-olympics-edition/>
- Bereby-Meyer, Y. and Grosskopf, B. (2008). Overcoming the Winner's Curse: An Adaptive Learning Perspective. *Journal of Behavioral Decision Making*, Vol. 21 No., 15-27.
- Cable, J., Henley, A. and Holland, K. (2002). Pot of Gold or Winner's Curse? An Event Study of the Auctions of 3G Mobile Telephone Licences in the UK. *Fiscal Studies*, Vol. 23 No. 4, 447-462.
- Casari, M., Jackson, C. and Zhang, J. (2009). Do Groups Fall Prey To The Winner's Curse? Department of Economics Working Papers, McMaster University.
- Chua, C. and Luk, P. (2005). Be a Winner Not a Loser: Experimental Evidence of Winner's curse. *The Marketing Review*, Vol. 5 No. 1, 303-314.
- Feltovich, N. (2006). Slow Learning in the Market for Lemons: A Note on Reinforcement Learning and the Winner's Circle. In J. Lakhmi, C. Chung, & H. Shu, *Computational Economics: A Persepective from Computational Intelligence* (pp. 149-160). London: Idea Group Publishing.
- Holt, C. A. and Sherman, R. (1994). The Loser's Curse. *The American Economic Review*, Vol. 84 No. 3, 642-652.
- Idson, L.C., Chugh, D., Bereby-Meyer, Y., Moran, S. and Grosskopf, B. (2004). Overcoming Focusing Failures in Competitive Environments. *Journal of Behavioral Decision Making*, Vol. 17 No. 3, 159-172.
- Kagel, J. L. and Roth, A. E. (1995). *The Handbook of Experimental Economics*. NJ: Princeton.
- Kambil, A. and van Heck, E. (2000). Information Leverage in Auctions and Markets. *Accenture: Institute for Strategic Change*.
- Kaplan, D. and Glass, L. (1995). *Understanding Nonlinear Dynamics*. New York: Springer-Verlag.
- Salmon, T. C. (2001). An Evaluation of Econometric Models of Adaptive Learning. *Econometrica*, Vol. 69 No. 6, 1597-1628.
- Samuelson, W. F. and Bazerman, M. H. (1985). The Winner's Curse in Bilateral Negotiations. *Research in Experimental Economics*, Vol. 3 No. 1, 105-137.
- Savikhin, A., Maciejewski, R. and Ebert, D. (2006). Applied Visual Analytics for Economic Decision-Making. *Working Paper*.
- Selten, R., Abbink, K, and Cox, R. (2005). Learning Direction Theory and the Winner's Curse. *Experimental Economics*, Vol. 8 No.1, 5-20.
- Thaler, R. H. (1988). Anomalies: The Winner's Curse. *Journal of Economic Perspectives*, Vol. 2 No. 1, 191-202.
- Vega-Redondo, F. (1997). The Evolution of Walrasian Behavior. *Econometrica*, Vol. 65 No. 2, 375-384.
- Warren, H. and Samuel, S. (2009). The Winner's Curse and Optimal Auction Bidding Strategies: Using Management Science Tools to Develop Practical Strategies in Competitive Bidding Environments. *Graziadio Business Review*, Vol. 12 No. 2.

The Limits of Friendship

Alex M. Verbeek



ABSTRACT

The purpose of this thesis is to contribute to contemporary philosophical debates on friendship. These debates focus primarily on either capturing the nature and value of friendships (both to the self and society), or on determining the extent to which friendship and its attendant duties challenge moral theories that hold impartiality as their highest ideal (primarily consequentialist and deontological moral theories). In the literature on friendship there has been surprisingly little overlap between these two debates: accounts of the characteristics of friendship tend to ignore – explicitly or not – issues of morality; conversely, philosophers engaged in justifying the demands of friendship with regard to ethical demands allow their ethical commitments to taint, and generally over-simplify, their conception of friendship.

This article seeks to unify these two debates by exploring the limits friendship imposes on itself rather than marking out the limits of friendship as determined by an abstract ethical theory. It does so by recognizing an important challenge made to them by Stroud (2006) who argues that friendship is characterized by, and indeed requires differential doxastic and epistemic practices. This is an important consideration for accounts describing the internal qualities of friendship, but more importantly it poses a serious challenge to impartialists who argue that friendship is justified within – and so compatible with – moral and epistemic impartiality. By incorporating this observation into the characterization of friendship, this thesis advances a more subtle interplay between friendship and morality by scrutinizing the limits of this epistemic bias with regard to one's friends.

THE LIMITS OF FRIENDSHIP

Friendship has long been held as one of the central goods in our lives, and is defined in the Stanford Encyclopaedia of Philosophy as 'a distinctively personal relationship that is grounded in a concern on the part of each friend for the welfare of the other, for the other's sake, and that involves some degree of intimacy' (Helm "Friendship"). The canonical theory of the value – and indeed virtue – of friendship is to be found in Aristotle's *Nicomachean Ethics*. Here Aristotle distinguishes between three kinds of friendship, which are predicated on the reasons we have to form friendships: friendships of utility,

of pleasure and of virtue (*Nicomachean Ethics*, Book VIII). Most discussions about the nature of friendship (whether from an Aristotelian position or not) are based on these three distinctions. For Aristotle friendship – especially between virtuous persons – was closely linked to living a virtuous life. In this sense one could say that the nature of friendship – what effect it has on the persons involved and the actions that characterize it, or are required by it – for Aristotle overlaps significantly with the justification of friendship. The justification for friendship is that it simply is necessary for a happy, flourishing and virtuous life.

The special, even essential, value attributed to friendship (and derived from Aristotle's theory) has led to it becoming an important consideration in moral theory because it highlights the tension of the impartialist-partialist debate. Impartiality here is defined as 'the absence of bias or favouritism toward oneself or one's own, whether one's own projects, loved ones, or just those related to one in some way felt to be significant' (Vice 294). Friendship, on the other hand, seems to demand partial concern and treatment, and, furthermore, that actions undertaken out of such partiality are valuable in their own right. Not only do we do things for friends that we would not as soon do for complete strangers, our commonsense conception of friendship tells us that we should do things that we would not do for strangers because they are persons of special concern to us. In other words, it seems constitutive of friendship that our friends do things for us because they are our friends, not out of moral considerations for the greater good.

The 'moral pull' of friendship is such that we would not accept a moral theory that did not 'make space' or recognize both the value friendship has to us as individuals and its social value. As a consequence of this, much of the recent work on friendship has been on justifying the special concern we feel for particular persons (and the subsequent duties we seem to have to our friends) in the face of the intuition most of us have that morality requires of us that we treat everyone impartially. Responses to this range from claims that the impartiality thesis is untenable (Cottingham 1981; Cottingham 1986), to declarations that the demands of impartiality are often overstated and are in fact compatible with friendship – if not actually encourage it (Baron 1991; Bernstein 2007; Jeske 1997), and that the debate between partialists and impartialists is actually one between the demands of love and the demands of morality (Wolf 1992).

What these accounts have in common is that they give a positive account of friendship with regard to the impartiality thesis – that is to say, they all

describe when the demands of impartiality (if they hold it at all) should be contravened in favour of the demands of friendship. Put simply they justify friendship. Similarly, the accounts of the nature of friendship focus primarily on the positive effects such relationships have on us – or, to put it precisely, give an account of the effects of personal relationships in positive terms (why they are of higher value to us than our relationship with strangers). While this might seem to be a cynical point to raise as virtually no one would contend that friendship is not valuable to us, it does have an effect on how the partialist-impartialist debate is structured. Bearing in mind the innate and irrefutable, if somewhat difficult to define 'goodness'¹ of friendship I propose to challenge the debate by (essentially) taking friendship as justified. By this I mean that I shall explore the value and duties of friendship 'on its own terms' – i.e. by leaving them unprejudiced by the normative considerations of the need to retain some predetermined level of impartiality. Stroud (2006) argues that friendship demands what mainstream epistemological theories call 'epistemic irrationality' that contradicts the ideal 'epistemic rationality' that modern ethical theories – those that advocate impartiality – assume and indeed require. What I propose, then, is to take seriously the consequences and demands of this 'epistemic irrationality' and to explore the limits it imposes on itself. This move removes much of the implicit guilt in the impartialist theories (no matter how it is phrased, or how much its advocates profess this not to be the case, the justification of friendship against the backdrop of the impartiality leaves one feeling that one must always be on guard against giving an undue amount of attention to your friends) as it fully embraces the 'moral goodness' of both friendship and moral impartiality. Rather than suggest more justifications for forming and developing friendships (both for the value it 'intrinsically' has for ourselves qua individuals, and against the impartiality thesis) I shall explore when one is justified (morally) in ending

a close friendship. By ending a close friendship I mean that one has extended one's interpretive charity with regard to your friend's revealed character or action to its absolute limit. It is the point at which one can no longer, in good conscience, accept some part of your friend's character; the point at which your friend's actions or beliefs violate your moral core to the extent that you are willing to forsake the commitment to each other (and feel justified in doing so) that friendship entails. It is my hope that this shift of optics in the debate will yield a sharper characterization of the nature of friendship, its value (both individual and social) and its demands or duties.

I. THE NATURE OF FRIENDSHIP

Before sounding out the limits of friendship I will first conduct an internal investigation of friendship in order to explicate a credible account of this particular relation. This is important because, as Cocking & Kennett state, 'a plausible account of that nature of friendship would seem not only helpful but necessary to ground and offer direction to discussions about the value of friendship' ("Friendship and the Self" 502). While it is customary to follow Aristotle's canonical analysis of friendship, the account advanced in this chapter will primarily be drawn from Cocking & Kennett's "Friendship and the Self" (1998) and Stroud's "Epistemic Partiality" (2006). The reason for this is that for Aristotle friendship – especially 'ideal' friendship between two morally virtuous individuals – is closely linked to living a virtuous life. In this sense one could say that the nature of friendship – what effect it has on the persons involved and the actions that characterize it (or, indeed, are required of it) – for Aristotle overlaps significantly with the value and justification of friendship. However, as the quote from Cocking & Kennett's paper illustrates, one should not conflate the nature of friendship with its value (either to us as individuals or to society in general). The reason I follow Stroud's, and Cocking & Kennett's characterization of friendship, then, is because they are not concerned with developing an account of

friendship with regard to a particular moral theory.

A few more words before we begin. Although I am – at this point – not developing an account of friendship with regard to a particular moral theory, I will highlight particular aspects of friendship that have the greatest bearing on my research question. With this in mind, I shall not attempt to give an exhaustive account of friendship as this is beyond the scope of this project. Although both the articles by Kennett & Cocking and by Stroud are concerned with the nature of friendship, they are still concerned with particular aspects of it: Kennett & Cocking are concerned with the ways (close) friendship has an impact on the self, while Stroud is concerned with the proper epistemic bias of the good friend. As the purpose of this thesis is to explore the limits of friendship I believe I am justified in similarly limiting my treatment of the nature of friendship.² Having said this, I shall give a very brief overview of the debate over the nature of friendship, as it will help contextualize the two papers I am drawing on.

CONTEXTUALIZING THE DEBATE ON FRIENDSHIP

In the entry on friendship in the Stanford Encyclopaedia of Philosophy (SEP) Helm (2009) defines friendship as:

[A] distinctively personal relationship that is grounded in a concern in part of each friend for the welfare of the other, for the other's sake, and that involves some degree of intimacy. As such, friendship is undoubtedly central to our lives, in part because the special concern we have for our friends must have a place within a broader set of concerns, including moral concerns, and in part because our friends can help shape who we are as persons (Helm "Friendship").

Helm points out that the main themes that recur in debates of philosophical accounts of the nature of friendship are: 'mutual caring (or love), intimacy, and shared activity' (ibid.).³ Although the emphasis on these elements may vary in different accounts

¹ At the risk of inconsistency, I do not equate – or limit – the concept of the 'essential goodness' of friendship with the (neo)-Aristotelian idea that friendship is a 'good in itself' insofar as two virtuous persons recognize (and develop) some innate moral character in each other. This is not to say that (close) friendships are not based on character, or on esteem for each other's merits (in a broad sense that is not limited to moral merits). Friendship, here, is good qua a relationship not qua a relationship between morally virtuous persons.

² I wish to make another methodological remark at this point. In this section I first discuss Cocking and Kennett's paper, and then move on to Stroud's analysis of friendship. I do this because, as stated before, Cocking and Kennett are concerned only with how friendship impacts the self while Stroud draws out aspects of friendship by looking at how we would (and should) act, qua friend, in interactions with third parties.

of friendship, they are all intertwined to create a complete picture of friendship. What has received the most attention (especially in light of conflicts with moral theories) is the basis of the mutual caring within friendship. The majority of moral theorists follow Aristotle in claiming that ideal friendships are those in which friends care for the other for the other's sake (i.e. for who they are as opposed to, for example, their usefulness to you).⁴ However, while for Aristotle this meant caring for each other through love for each other's virtuous or moral character, many contemporary philosophers depart from his theory here. It is, therefore, important to point out here that that the type of friendship that Stroud and Cocking & Kennett discuss is roughly what Aristotle classifies as friendships of virtue, the highest of the types of friendship in his *Nicomachean Ethics*. While Aristotle bases virtue friendships (roughly stated) on the love of each other's shared moral character, Stroud and Cocking & Kennett discuss friendships that are not (necessarily) based on the individual's *moral* character in order to develop a 'morally neutral' account of friendship. Cocking & Kennett state that their investigation concerns only 'those intimate friendships which feature reciprocal deep affection, well-wishing, and the desire for shared experiences' ("Friendship and the Self" 502). Similarly, one of the two background assumptions that Stroud bases her thesis on is that 'friendship is in some importance sense based on your friend's character and on esteem for his merits' ("Epistemic Partiality" 511), while adding that this need not narrowly mean moral merits.

Although it may seem somewhat premature to mention the second of Stroud's two background assumptions,⁵ it will help clarify the disagreement about the basis for the mutual caring and intimacy in friendship and frame the debate about the tension between friendship and (impartial) morality. This second assumption is this:

[F]riendship is or involves a kind of commitment...[and] having commitments means not deciding every issue solely on its merits, if we mean by that dispassionately adjudicating and issue "from nowhere" (ibid. 511).

Put differently, being in a friendship involves not requiring your friend to constantly prove anew that he is a good person (i.e. a person worthy of your friendship) according to a set of pre-determined standards. To do so would give your friendship an unpleasantly contingent flavour. Instead, the commitment friendship involves assuming (or at the very least taking into consideration) the good character of your friend when confronting and deliberating new situations involving them. Stroud's characterization of disinterested, impartial adjudication as coming 'from nowhere' makes clear that our commitments to our friends come 'from somewhere.' This 'somewhere' is the mutual caring and intimacy that constitute friendship. In other words, intimate friendships produce their own commitments – and it is the mechanisms by which this is done that Stroud and Cocking & Kennett investigate.

There are various accounts of mutual caring

and intimacy that are constitutive of friendship. In his entry on friendship Helm states that, 'a central difference among the various accounts of mutual caring is the way in which these accounts understand the kind of value evaluation implicit therein' (Helm "Friendship"). The axis along which these accounts are divided is whether the value evaluation is a matter of appraisal or of bestowal. In broad strokes, the disagreement concerns whether we value our friends because of the (good) character we discover them to have (appraisal), or whether we project an intrinsic value onto our friends (bestowal). The former account is more common and is the one that both Stroud and Cocking & Kennett subscribe to. One reason why the bestowal account is thought to be attractive is because it is seen to be the one that explains why our friends are not fungible. The argument is that if the basis of our caring for our friends is based on our appraisal for his good character, then we do not value our friend intrinsically because someone else might equally have these character traits, thereby allowing us to switch friends without any real loss. This conclusion, however, is pre-emptive because the reasons we have for entering friendships are different from those for which we have for continuing them. While we might enter friendships because we value the character of our friends, this need not entail that once the friend's character changes (which is precisely a consequence of the profound influence friends have upon each other) the reason for the friendship disappears. Instead, 'through the friendship, and through changes in your friend over time, you may come to change your evaluative outlook, thereby in effect subordinating your commitment to certain values to your commitment to your friend' (Helm "Friendship"). In essence, this is the same argument that Stroud makes in her second background assumption about the commitment that is characteristic of friendship. Since the reciprocal nature of friendship is described extensively by Cocking & Kennett I shall now turn to their account of friendship. The point to keep in mind here is simply that caring for your friend for his sake and basing this

on an appraisal for his good character in the manner discussed above is compatible even if your friend's character and values (and your own) are changed by your friendship.

THE FLUIDITY OF THE SELF IN FRIENDSHIP

In "Friendship and the Self" Cocking & Kennett are concerned with how (accounts of) friendship view 'the self in friendship and the role they give to disclosure of the self in the construction of intimacy' ("Friendship and the Self" 502). The two accounts they consider, and reject, are what they call the 'mirror view' and the 'secrets view' of friendship. The mirror view holds that intimate friendships are marked by the great extent to which we see ourselves reflected in our friends; the secrets view holds that companion friendship is marked by the great extent to which we disclose ourselves to our close friends. Cocking & Kennett argue that, while the two accounts point to different phenomena, they are 'united in holding that central to the trust and intimacy in companion friendship is that one's self is disclosed in the relationship – either I disclose my self to the other or my self is disclosed in the other' ("Friendship and the Self" 503). They claim that these two accounts fail to capture both the features of close friendship and the distinctive ways in which they have an impact on the self.

In their stead Cocking & Kennett propose their drawing account of friendship, which holds that: 'as a close friend of another, one is characteristically and distinctively receptive to being directed and interpreted and so in these ways drawn by the other' (ibid.). By 'being directed' they mean that it is characteristic of companion friendship that we value our friend's interests (and beliefs, values, etc.) simply because they are our friend's. For example, I might develop an interest in skiing solely by virtue of my friend's interest in skiing.⁶ Put simply, we are generally more given to be receptive to our friend's interests, and to expend more energy trying to understand why our friend has these interests (and beliefs, values, etc.).⁷ While it is often an effect of this

³ Admittedly, the importance of shared activity is not featured in the quote in part because it is often taken for granted, and also because it is dependent on how a particular account of friendship construes the type of intimacy found in friendship. The account of friendship given here will also focus primarily on the aspects of mutual caring and intimacy.

⁴ In Aristotelian terms, ideal friendships are called 'virtue-friendships.' The other two types discussed in his *Nicomachean Ethics* are friendships of utility and of pleasure, which are seen to be less valuable because these friendships are contingent on the continued usefulness or pleasure you derive from your friend. In other words, it is commonly held that in these friendships you do not care for your friend because of who he is. However, for a criticism of this view see: Grunebaum, James. "Fair-Weather Friendships." *The Journal of Value Inquiry* 39 (2005): 203-214.

⁵ At the end of her internal characterization of friendship – specifically the epistemic bias it involves – Stroud offers two background assumptions she makes to offer an account of friendship within which her claims about the proper epistemic bias of friendship are plausible. See pp. 510 – 512.

⁶ This might seem a trivial example given the subject of the paper. While it does lack a certain 'moral gravity,' I chose a more everyday example because this characterization is supposed to leave aside moral questions. Given this, however, it is not hard to see how this process can manifest itself around more morally serious issues. I might have a friend who is passionate about sustainable living and I might, because of her passion, be convinced to be more conscious of how I live on this planet.

mutual responsiveness that friends grow to be more similar, it is this responsiveness (and not the effect thereof) that characterizes companion friendship. The second significant feature of the drawing account of friendship – that our friends interpret us – highlights the process through which we not only are made aware of our character traits, but also that ‘the close friend’s interpretation of the character trait or foible can have an impact on how that character trait continues to be realized’ (“Friendship and the Self” 505). It is thus a constitutive feature of close friendship that we interpret our friend’s character, and that we are likewise open to being interpreted by our friends. The result of this willingness to be interpreted is that our character and self-conception is partially shaped by our friend’s interpretation. In short, ‘[on] a drawing account the self is conceived as a relational thing that is, in part, developed or moulded through the friendship, and this process of mutual drawing seems...central to the establishment of the intimacy of companion friendship’ (“Friendship and the Self” 505-506). Cocking and Kennett sharpen their drawing account of friendship by contrasting it with the mirror and secrets views of friendship in greater depth. I shall follow Cocking & Kennett’s analysis and highlight salient features of this discussion in order to provide a nuanced picture of friendship.

The mirror view of friendship is derived from the Aristotelian conception of virtue friendship, which holds that friendship is based on the mutual

recognition of ‘pre-existing “firm and stable” features of another’s character [i.e. virtues]’ (“Friendship and the Self” 506). As well as being ‘firm and stable,’ these features are recognizable to one another because they are shared (i.e. friends have them in common), whereby friendship becomes a form of proper self-love based on the appreciation of similarity. As mentioned earlier, Cocking & Kennett waive the claim that close friendship is open only to the virtuous and focus instead on the claim that mutual acknowledgement of similarity is important. There are several things at stake with this claim. First of all, it is unclear how the development towards intimate friendship could be marked by recognition of greater resemblance between two individuals.

The discovery that a fellow student also enjoys skateboarding does not, in and of itself, direct me to spend more time with him, nor move me to attempt to develop an intimate friendship with him.⁸ It often happens, however, that we are directed to spend more time with our friends doing things they are interested in simply because they are our friend’s interests (i.e. for our friend’s sake). This suggests that the development of companion friendship need not even be helped by, let alone be conditional on, the discovery of shared tastes or interests. The observation that difference can be important (and even beneficial) for friendship is one that carries through to other, deeper aspects of friendship.

On the Aristotelian account of friendship the mutual recognition of our selves in each other lead friends to love each other as a natural extension of self-love.⁹ It is through this mutual recognition that ‘my self-image is confirmed through seeing myself reflected in you’ (“Friendship and the Self” 509). Briefly put, the objection to this account is that it ascribes a passive and shallow role to me qua friend, and that it has a static conception of the self. As a friend I do not passively, nor objectively, reflect my close friend’s characteristics, but instead I reflect back my interpretation of my friend. For this I do not need to be like my friend; in fact, our relationship might be enriched by the fact that we are not alike, nor should I expect that my close friends all have the same interpretation of me. Cocking & Kennett argue that a better – though still limited¹⁰ – analogy of friendship than the mirror is that of the portrait painter: just as the portrait painter draws his subject in a new light ‘so, too, do our close friends draw us and so enrich our sense of self through their engaged

interpretations of us...We are thus, to some significant extent, each other’s creators’ (“Friendship and the Self” 509). Through this process we are moved to see aspects of our own character through our friend’s interpretation because we value their interpretations. This can have a significant bearing on how we see ourselves, and on how these characteristics manifest themselves.

This dynamic aspect of friendship points to another problem with the mirror view of friendship. Specifically, it ‘posits or implies a mistaken view of the self or of the self in friendship...[namely] that we come to friendship as fully formed and self-sufficient individuals’ (“Friendship and the Self” 509). The drawing account, on the other hand, states that through the process of mutual interpretation the self is partially constructed through our separate friendships.¹¹ Cocking & Kennett present two interpretations of the mirror view that attempt to overcome this problem.¹² The ‘objective mirror view’ argues that intimate friends (in spite of apparent

⁷At this point I wish to introduce two caveats that Cocking and Kennett introduce themselves: that we often develop interests ‘simply’ because they are my friends does not (necessarily) mean that I will be directed by my friend if I find some of his pursuits inane or deplorable. However, ‘the fact that there may be constraints on the interest one can have in one’s friend’s interests does not touch the point we make here. It remains true to say that one can be disposed to be interested in pursuing certain activities that one otherwise would not be, simply on account of one’s friend’ (Cocking & Kennett 504). In a similar vein the term ‘directed’ should not be taken to imply an inherent connotation of domination and control in the friendship – each friend is as likely to be directed as the other (ibid. 504-505).

⁸Although we are now ignoring virtues of character, it is also hard to see how the discovery that a fellow student is generous or honest (on the condition that I am as well) will direct me to become friends with her. Cocking & Kennett augment this argument by stating: ‘I may of course be disposed from motives other than friendship to pursue activities that we have in common. I may be motivated by ambition to spend more time with you. What this also suggests, then, is that even if the recognition of greater similarity between us provides the occasion for us to spend more time together this need not indicate a move toward companion friendship’ (“Friendship and the Self” 508).

⁹ Cocking & Kennett: ‘It follows then that I cannot love you for any characteristics which I am unable to love in myself, and it is this kind of thought that underpins Aristotle’s view that companion friendship is only available to the virtuous, for only they can truly love themselves. There is some truth in this last claim. Perhaps I am not inclined toward friendship with you because I do not like what I see of me when I look at you. I see my mean streak reflected in you or my tendency to brood over imagined slights. I am surely more likely to be inclined toward friendship with someone in whom I see reflected my particular sense of humor or my burning concern with social issues’ (508).

¹⁰ ‘And here we note an important difference between the interpretations of the portrait painter and that of the friend. As is well recognized, friendship is a reciprocal relationship. The reciprocity of the relationship itself influences the process and the outcome of creative interpretation in friendship. This is not true of creative interpretation in the portrait painter case. The subject is passive with respect to the interpretation’ (Cocking & Kennett 509).

¹¹ What follows from this, but is not stated by Cocking & Kennett explicitly, is that different close friends might have ‘dissenting’ interpretations of me, and so it follows that I have ‘different selves’ within each of these relationships. While this may have been overlooked I do not think that this need worry us. I think it is a common phenomenon that we feel we ‘act differently’ (or feel like a slightly different person) among our different close friends (and other people as well). Following Cocking & Kennett this is an effect of the different interpretations our friends have of us (some might value certain character-traits more than others, depending what they are like), the strength of which on our self-conception is dependent on how much we value their interpretation. Anyway, while there may be differences these are likely small – it is unlikely that one close friend will find you the epitome of generosity while another finds you selfish.

¹² For the full analysis of both accounts c.f. “Friendship & The Self” pp. 510-514.

differences of character) have an intuition of a deeper, shared character and by studying each other come to realize existing and objective, but previously unknown, character traits (for example, because of John's love of opera I might discover a latent passion for opera). There are several problems with this account: one is that it is not clear why your similarity to an other increases your self-awareness, but not dissimilarity – which seems at least as plausible (if not more so); another is that friends alter our self-conception not only by making unknown character traits visible to me, but also by (re)interpreting character traits I am already aware of.¹³ This last point demonstrates that the 'objective mirror view' does not solve the problem of the passive role qua friend as on this account friends interpret themselves through the other rather than each other. Nancy Sherman's 'ideal mirror view' – which posits that friends function like an ideal mirror because they 'present to us a more ideal self' ("Friendship and the Self" 513) – fares little better. This contention need not be limited to the idea that friends provide an ideal image to which we should aspire, but can play a more active role in interpreting our character. There are, however, still two problems with this view: firstly, 'even if this ideal mirror view allows for some active interpretation by the friend, it does not seem an essential feature of friendship that this will have the result that one becomes more like one's friend' (Cocking & Kennett 514);¹⁴ and secondly, the claim that friends provide a moral ideal is a highly moralized account of friendship that is out of touch with everyday experience.¹⁵ The mirror view (even generously interpreted), then, does not capture the nature of friendship with the depth and subtlety of the drawing account and thus the latter should be

favoured over the former.

The second account of friendship that Cocking & Kennett consider is the 'secrets view' of friendship. This view is characterized by the idea that friendship is brought about by self-disclosure to the other (i.e. telling secrets or private information). The reasoning behind this view is that the disclosure of intimate information (primarily regarding motives for certain actions) puts us in a vulnerable position, which shows trust and establishes intimacy. Furthermore, it gives our friends privileged information about ourselves so that they are able to comment on our lives (generally give advice), thereby helping us flourish. The secrets view of friendship holds that without the self-disclosure of secrets, one cannot claim to be close friends. While this view is more plausible than the mirror view of friendship – especially 'the idea that companion friends have a commanding or privileged perspective on each other's lives' ("Friendship and the Self" 515) – there are still some concerns that should be raised. First of all, it is not clear there is a straightforward causal link between the disclosure of secret information and the development of intimacy that is necessary for close friendship. I might, for example, disclose personal information to my therapist, but it is clear that we will not become friends as a result. This is not to deny that the sharing of secrets is unimportant in friendship, but it does not account for the commanding perspective of each other's lives. Instead, it is important that friends tell each other what they value that accounts for this privileged perspective. It is not the sharing of secrets that determines the strength (or intimacy, value, etc.) of a friendship, but rather the friendship determines what type of information we choose to share: 'it is the

value we assign to the hopes and concerns we share with each other (whether we wish them to be kept private or not) and the fact that we choose to talk to each other about what matters to us that contributes to the growth of intimacy between us' (Cocking & Kennett 518). The drawing account of friendship can explain this because the mutual drawing and interpretation it describes 'sheds light on how the shared valuing that goes on in friendship, and the intimacy that comes with this, are guided and shaped within friendship' (ibid.). Having established why both the mirror and secrets view should be rejected in favour of the drawing account, I now turn to what more the drawing account reveals about friendship.

THE GOVERNING CONDITIONS OF FRIENDSHIP

To gain a sharper picture of what distinguishes friendships from other personal relationships, and to show how the drawing account of friendship offers a convincing account of this, Cocking and Kennett analyse the governing conditions of friendship. These are 'accepting conditions' – conditions under which different kinds of relationships are entered into and developed – and 'terminating conditions' – conditions that lead to the ending of this relationships. Given the importance of direction and interpretation in friendship, for Cocking and Kennett it is an acceptance condition of friendship that we are open to being directed and interpreted. That is to say, under the drawing account of friendship we could not say that a person with an excessively rigid personality – one who is not open to being directed and interpreted by others, and doing so for others – could ever be a close friend to anyone.

For the purposes of this paper the terminating conditions are more interesting to consider. On the drawing account of friendship the terminating condition of friendship is the converse of the acceptance condition: it is when a person 'can no longer accept or engage in direction and interpretation with one another' ("Friendship and the Self" 521). Cocking & Kennett point out that a common – and commonly seen as sufficient – reason that is given in explanation for the end of a friendship is the loss of anything (significant) in common with the friend. However, as demonstrated by the criticism of the mirror view close friendships need not be precluded by dissimilarity – indeed, friends who see each other again after many years (and so have 'grown apart')

may have changed significantly but are still willing to be directed and interpreted by one another. On the other hand, 'it seems just the right thing to say, of those old friends whose interpretations of me I can no longer accept, that we have grown apart' ("Friendship and the Self" 521). While it is true that in our lives these cases are ones in which there is a waning of influence between two friends for which there is no one specific reason, they do point us in the right direction of why we would consciously choose to end a close friendship. Intuition tells us that there is a point at which our friends violate our sense of self – a self that was in fact partially created by that particular friendship – to the extent that we can no longer bring ourselves to be interpreted by them and thus end that friendship.

This summary concludes Cocking & Kennett's account of the nature of attachment and intimacy in friendship, which explains why we are '(partly) determined by the friend's interpretation of me and [why we] have reasons to act that are directed by [his or her] interest' ("Friendship and the Self" 527). While Cocking & Kennett do consider why we have reasons to act that are determined by our friend's interests, they do not consider how this might manifest itself with regard to others. That is to say, Cocking & Kennett only consider direct mutual interpretation (sharing friends' interests, etc.) but not what we might call indirect interpretation. These are cases in which we (are forced to) interpret our friends based on their actions with third parties, or at least reports thereof. This is important to consider because friendship is never a sealed relationship and especially because these considerations (i.e. how our friends treat other people) often figure in our deliberations of whether to end our friendship. For a discussion on this I now turn to Stroud's "Epistemic Partiality in Friendship."

FRIENDSHIP & BIASED BELIEF

The internal investigation of friendship that Stroud advances in "Epistemic Partiality in Friendship" complements – indeed strengthens – Cocking and Kennett's drawing account of friendship. Stroud argues that, 'friendship places demands not just on our feelings or our motivations but our beliefs and our methods of forming beliefs' ("Epistemic Partiality" 499). As a result of this the doxastic practices particular to friendship engender a skewed set of beliefs toward our friends. It is not hard to see how

¹³ 'The evaluative, creative content in the interpretations offered by my friend makes them dynamic; they will often change the way I view myself and the way in which some character trait of mine is realized, even when the trait in question is already known to me' (Cocking & Kennett 512-513). Cocking & Kennett illustrate this point by providing an analogy of seeing your nose in a mirror. If you see your nose in a mirror for 100 times you are unlikely to change your opinion of it, but if friend tells you your nose is handsome you will (likely) see it differently from then on.

¹⁴ This is not to say it never happens, just that it does not always happen.

¹⁵ Indeed, if this were the case this paper would have no purpose. For a considered account of the ways in which friends can lead us 'morally astray' see: Cocking, Dean and Jeanette Kennett. "Friendship and Moral Danger." *Journal of Philosophy* 97 (2000): 278-296

such a bias arises from the same basis of friendship as the drawing account of friendship: since we value our close friend and the relation of friendship we have with him (and so also on the degree to which we allow ourselves to be directed and interpreted by him) we also have an interest in preserving the value we see in that person.

Stroud's phenomenological analysis of friendship follows and builds upon Michael Stocker's "The Schizophrenia of Modern Ethical Theories" (1976). One of Stocker's main claims in this paper is that the friend's motives are an important component of friendship as well as his actions (i.e. it is important that friends do things for each other out of friendship rather than, for example, moral duty). While Stroud does acknowledge that Stocker uses this argument to show that these motives are precluded by modern ethical theories, she puts aside these considerations and argues that what is important in his paper 'is that there is a distinctive moral psychology associated with friendship' ("Epistemic Partiality" 501).¹⁶ Stroud's methodology also follows Stocker's: namely that if we have an intuition that a good friend would or would not do something, then we assume that being a good friend involves doing or not doing this. That is to say, this intuition 'becomes a constitutive element or feature of friendship; or, finally, that this is a demand of friendship: something you owe your friend' ("Epistemic Partiality" 501 – 502). There are two qualifications to this: first, that the demands of

friendship are not to be understood in an exceedingly strict sense – failing one's 'duty' to one's friend once (or from time to time) does not disqualify you from being a good friend; secondly, the demands of friendship are not necessarily moral demands.¹⁷

With this methodology 'on the table' we can now engage more thoroughly with Stroud's claim that 'friendship also places distinctive demands on our beliefs and our belief-forming procedures' ("Epistemic Partiality" 502-503). The way Stroud elucidates these distinctive demands is to demonstrate how they manifest themselves in the public realm – for example, when you hear something (mildly) bad about your friend from someone else. The example Stroud uses is a case in which you hear that your friend acted disreputably and you do not know that what you hear is false. In such cases you have a duty qua friend to stick up for your friend – to publicly defend his reputation. But, as Stroud argues, a little introspection reveals that our duty as a good friend demands more than this: a good friend will react differently, and form different beliefs, upon hearing this story than would a stranger or detached observer. As Stroud puts it: 'our ways of dealing with evidence seem systematically to shift when that evidence bears on the actions and characters of our friends: we seem in a variety of respects to adopt differential epistemic practices when our friends are at issue' ("Epistemic Partiality" 504). Again, the argument is that not only do we do this, but also that it fits our conception of what

a good friend should do. While this argumentation will become more salient later, it is important to stress that here we see that within our personal conception of what friendship is there is a reflection to what friendship involves (that is, to friendship from what one might call a 'social perspective'). In many respects these differential doxastic practices have a similar texture to Cocking and Kennett's claim that it is a constitutive feature of close friendship that there is mutual interpretation of character – the difference is that Stroud extends this claim into the public domain and points out that in this domain the interpretation is generally favourable.

Stroud identifies two broad categories into which the differential responses manifest themselves. The first 'locus of difference is the cognitive activities we engage in when processing new data about our friends. What is distinctive in this domain is that we tend to devote more energy to defeating or minimizing the impact of unfavourable data than we otherwise would' ("Epistemic Partiality" 505). It is thus constitutive of friendship that when we hear bad things about the friends we devote more energy (and resources) to scrutinizing the credibility of the evidence being presented to us. For example, we might question whether the person telling the story has ulterior motives for doing so, or ask ourselves whether the person in question is not given to exaggerating such stories. In addition to this we also assign a greater degree of plausibility to alternative interpretations of what we hear than we would for someone who is not our friend.

The second 'locus of difference' is that, in addition to the differential methods of interpretation, we end up drawing different conclusions than we otherwise would (or a detached observer would) with regard to our friends. In other words, not only do we come up with alternative explanations of the reported conduct, we also are more likely to believe these explanations than we would of a nonfriend. This is not to say, as Stroud is careful to stress, that we would flatly deny incontrovertible facts about our friend's conduct. Instead this interpretive charity is a result

of the fact that people's motivations, character and actions are not transparent and thus always 'can be seen in different lights, put in different perspectives, filed under different labels and concepts' (Stroud 507). Furthermore, not only can we interpret people's character traits (and actions) differently, we can also, once interpreted, decide how important these traits are in our overall picture of our friend. In short, 'the bias of the good friend will normally take the form of casting what she sees or hears in a different light, shading it differently, placing it in a different optic, embedding it in a different overall portrait of her friend' ("Epistemic Partiality" 508). The exact point at which this bias begins or ends varies from case to case, and from person to person.¹⁸

Taking together the two accounts of the nature of friendship, then, we have a thorough account of how we are party drawn by our friendships (how the intimacy of friendship manifests itself, and why we have reasons to value our friends), why we have reasons to act out of interest for our friend, and how these actions manifest themselves in the domain of belief. The next part of this thesis will consider cases in which all of these aspects of friendships are disrupted and stretched to their limits.

II. FRIENDSHIP, MORALITY, AND BELIEF

FRIENDSHIP AND MORAL OVERRIDINGNESS

Having established an account of friendship and its doxastic practices, I now turn to the discussion of where these features cause tension between other commitments that have a (reason-giving) claim on us. The most prominent one of these is the conflict between the demands of friendship and those of morality.

In "Friendship and Moral Danger" (2000) Cocking & Kennett argue it is the nature of friendship itself – as described by the drawing view – that leads to a conflict with (impartialist) morality. This is meant primarily as an argument against highly moralized

¹⁶ I must note here that I find Stroud's addition of the term 'moral' here somewhat problematic – especially as this is the section in which she conducts a 'purely' phenomenological analysis of friendship. The implication here is that by failing to live up to this distinctive psychology of friendship one is immoral rather than simply being a bad friend, and thus being a good friend becomes equated with being a moral person (which is a distinctively Aristotelian conception of friendship that she rejects earlier). Stroud's internal analysis of friendship would not lose any force – or any of its relevance – by dropping the term 'moral,' and so I ask the reader to disregard the aforementioned implications for the time being.

¹⁷ One final point about Stroud's caveats: her claim that the demands of friendship are not necessarily moral demands could be taken to imply that they often are moral demands – especially given the term 'demand.' This, however, would be a mistake, as Stroud herself states: 'I think the simple fact that friendship...is an interpersonal relation makes a wider range of vocabulary for its constitutive features appropriate, even if we are neutral on the moral status or force of those constitutive elements' ("Epistemic Partiality" 502). Perhaps another way to put this is that we need not think of the 'demands' of friendship as moral demands at all because the demand in question would have a prior claim to us qua demand of friendship – the fact that to an external observer the demand is also moral is often incidental to us.

¹⁸ Stroud states that the general procedure of producing such an 'empirical slant' generally starts with attempting to deny what you hear about your friend and advancing from there, until, as a last resort, you 'relegate your attribution of a character flaw to your friend to an obscure portrait of him' ("Epistemic Partiality" 509).

accounts of friendship (primarily Aristotelian virtue-friendships), which, they argue, have no real traction in our daily lives. On their drawing account of friendship it is still possible, and indeed it seems to happen often, that we are moved by our friend's vices and are sometimes required to do something immoral for the sake of our friend. On the highly moralized accounts of friendship – those in which friendship “is most notably seen as a vehicle for moral self-examination and character improvement” (“Friendship and Moral Danger” 278) – this is not possible, and, furthermore, such a friendship would have no real value. Cocking & Kennett deny this charge and claim the converse is true – that a friend who only develops a friendship with you insofar this is acceptable within a given moral framework, cannot be said to be a true friend. Building on this they argue that: ‘Given the clear disvalue to friendship of conducting our friendships entirely within the moral framework, it may well not be reasonable to accept that whenever there is a conflict between friendship and morality, the moral considerations must be overriding’ (“Friendship and Moral Danger” 296). However, Cocking & Kennett introduce a caveat, which is that this argument does not extend to restricting cases such as committing a murder for your friend.¹⁹

It is not hard to agree with Cocking & Kennett that a friendship that sometimes leads you morally astray can still properly be called a friendship. The problem is rather with the phrasing of their claim about moral overridingness. They argue that because we sometimes are required by reasons arising from friendship to do things that are immoral (or at least, to do something other than what we have the most moral reason to do), and that friendship so conceived is still properly called friendship and still has value, then it is not reasonable to suggest that morality is always overriding. At first this seems to make sense, but the

conclusion Cocking & Kennett arrive at is strange: the claim that morality is not always overriding is paradoxical because a claim that is sometimes overriding is not really overriding at all. If Cocking & Kennett are right then it seems we pay a heavy price for friendship, for then we cannot expect friends to show restraint out of moral considerations (in interactions with nonfriends). That is to say, if morality is not overriding, then morality loses its claim on us as one could theoretically always point to one's friendship to claim that you had most reason to do an immoral act. This is why many moral theorists – which seem to include Cocking & Kennett in their claim that their argument does not extend to committing a murder for your friend – argue that moral reasons do (always) have an overriding claim on us. This suggests that a different, stable conception of moral overridingness is needed that is compatible with this more sophisticated account of friendship. The best way to do so is to explore the limits that friendship imposes on itself. For this I now turn to the analogous conflict Stroud sees between the demands of friendship and epistemology, and what this tells us about moral overridingness.

FRIENDSHIP & EPISTEMIC CONFLICTS

As I stated earlier in this essay, Stroud's claims about the epistemic partiality characteristic of friendship fits well with Cocking & Kennett's ‘drawing account’ of friendship. After concluding her internal investigation of friendship Stroud considers how this epistemic bias relates to mainstream epistemology (and, by extension, to morality). In order to draw out the features of these relations Stroud poses two questions: 1) how this characteristic bias would be evaluated by ‘mainstream contemporary epistemologies’ (“Epistemic Partiality” 512) and; 2) what we ought to do should we find a conflict between the epistemic demands of friendship and our epistemic ideals. It is the latter question that has

the greatest bearing on my project (and points to the conflict between impartialist morality and friendship that Cocking & Kennett see), but this cannot be properly addressed without first proposing some answers to the first.

In answer to the first question, Stroud claims that the doxastic practices of the friends constitute a genuine bias – ‘if bias is understood simply as differential epistemic treatment without epistemic justification’ (“Epistemic Partiality” 513) – and thus that these ‘distinctive epistemic practices would be unfavourably evaluated by standard epistemological theories’ (“Epistemic Partiality” 512). By ‘standard’ or ‘mainstream’ epistemological theories Stroud points primarily to evidentialist theories, which hold that you are only justified in forming a belief based on relevant evidence you have for that belief. That someone is your close friend is not in itself, according to these theories, a relevant epistemic reason to adopt the differential epistemic practices that we do. That is to say, there is nothing about the relation of friendship that warrants the good friend's doxastic bias. A second argument in favour of this position is that the beliefs the good friend arrives at when her friends are concerned ‘do not seem to be ones we could endorse from a purely epistemic point of view’ (“Epistemic Partiality” 513). Here Stroud is pointing to the fact that we tend to (and should) devote more energy to ‘explaining away’ – and believing our own explanations²⁰ – bad things we come to know about our friends than we do for nonfriends. In other words, rather than ‘being truth conducive, [the good friend's epistemic practices] seem to lead her into a distorted conception of reality’ (“Epistemic Partiality” 513). The doxastic practices, and the beliefs they engender, in the good friend are thus not justified according to the norms of the general standard criteria of epistemically responsible belief-formation. This is because, although the good friend does not regress into a complete denial of reality when bad evidence is presented to her, she nonetheless spends considerable resources searching for ways in which her belief in her good friend's character is justifiable in light of new (negative) evidence about her friend.

Here I would like to momentarily digress and rephrase this last point slightly by suggesting it is

more accurate to say that the good friend spends her resources for ways in which her commitment to her belief in the goodness of her friend's character is justifiable. This is a more nuanced way of phrasing the process as it shows that there is something at stake for her as well in continuing to believe the best of her friend. Earlier Stroud claims that friendship involves a type of commitment in which you do not judge your friend's merits anew every day. While I in no way claim that this is not the case, this does not mean that this commitment is unequivocal or unquestioning. If the drawing account of friendship is correct, then your friend's character (including his moral values) is partially shaped by you just as he shaped yours. It follows from this that your commitment to your friend not only says something about your friend but also about yourself (which, in turn, explains some of the motivation to believe the best of your friend). Friendship does not take place in isolation and in (publicly) standing up for your friend you also offer yourself up for judgement (albeit to a lesser extent because it is generally expected of friends to do so) by, as Stroud termed it, the court of public opinion. While this might, to some, seem to introduce an unpleasantly contingent flavour to friendship I do not think that this is the case. I would also like to point out something that Stroud failed to mention – that not only do we believe the best of our friends, but we generally also expect more or better of them than of nonfriends. So, another way of putting the limit of friendship is the point at which the friend violates your shared values, but also when your commitment to him would suggest that you endorse something that is at odds with your (moral) values.

Returning to “Epistemic Partiality,” Stroud claims that the epistemic practices demanded by close friendship run counter to the demands of mainstream epistemology. However, before turning to what we should do in light of this conflict Stroud forestalls a challenge which, if successful, could resolve the conflict before us by questioning whether the epistemological practices really are biased, or whether they run counter to general, impartial epistemological principles. This challenge is drawn from the ‘secrets view’ of friendship and argues that the friend is merely applying impartial, evidentialist

¹⁹ Of this example they write: “[I]f the pursuit of friendship within the moral framework meant that conflict occurred only where we were directed by the other in such base or extraordinary ways, then it would be reasonable to accept that we pursue our friendships only within our adopted morality. The problem presented by our cases, however, is that there is also a conflict between being governed by certain moral commitments and pursuing friendship in both admirable and perfectly ordinary ways’ (295).

²⁰ I shall return to this problematic point later.

practices more thoroughly than other people and that she is justified in forming different beliefs based on the large body of evidence she has about her friend in the capacity of being a friend. If this is correct, the friend is evidentially justified in being slower to draw a negative conclusion than others. This argument has a certain degree of plausibility: it is certainly true that we often have a 'privileged perspective' on our friend's motivations and character, from simply spending time together and indeed that we had a hand in directing the character of our friend, which allows us to reach a more accurate conclusion.²¹ Despite this plausibility, Stroud contends this argument does not wholly convince as the beliefs the good friend forms are not proportional to the evidence that is presented to her: 'she withholds belief in propositions amply warranted by the evidence at hand which would be natural inferences for a disinterested observer to make' ("Epistemic Partiality" 516).²² Furthermore, the appeal to the privileged epistemic perspective we have qua friend becomes less plausible once one takes into consideration that the greater prior knowledge you have of your friend has been shaped by this self-same bias.

To recapitulate the dilemma we are faced with, Stroud claims that, 'friendship requires epistemic irrationality' ("Epistemic Partiality" 518), and given the overwhelming reason we have to form friendship for their individual value we – by the same token – have very strong reasons not to live up to our epistemic

ideals. The question thus becomes what we ought to do with this dilemma. In response, Stroud argues that if friendship and the demands of our epistemic ideals conflict, and that we are forced to embrace something which these theories consider to be epistemic irrationality, then a solution might be to say that these theories could be wrong about what epistemic irrationality constitutes. To do this, Stroud argues, is to make a similar move that is available, and often made, at the discovery of a conflict between friendship and morality: 'For many moral theorists, the discovery of an incompatibility between morality and friendship would be proof, not that friendship is immoral, but that our previous ideas about morality were too narrow' ("Epistemic Partiality" 520). While this may seem to be the same response as Cocking & Kennett's claim that morality is not always overriding, this is not the case. Rather, Stroud advocates a change in our conception of morality in a way that preserves its authority, but allows for the practices of friendship.

The argument, then, is that friendship – due to its internal, personal value – can place similar constraints on formulating a plausible account of epistemic rationality. Thus, as in the moral case, 'rather than concluding that friendship is epistemically irrational, we could instead conclude that our previous ideas of epistemic rationality were too narrow' ("Epistemic Partiality" 522). Stroud anticipates that, more so than in the moral case, this

idea will be resisted because '[w]e don't think of the search for the correct standards of epistemic virtue as being in any way constrained by considerations having to do with a good life' (ibid.). That is to say, the standards of justifiable belief-formation are usually seen to be grounded in (objective) evidence, and not in agent-relative interests or pragmatic utility. Despite this hesitation to re-evaluate our standards of epistemic rationality, Stroud argues that we should be open to doing so. She holds given the doxastic practices of the good friend, that someone who wishes to maintain his belief in the epistemic 'ideal of purity' will 'have to relinquish any claim to overridingness for the epistemic standards he is busy formulating' ("Epistemic Partiality" 523). To do this, to concede that one's epistemic standards are 'not authoritative with respect to what we ought to believe all things considered...is a nontrivial price to pay: many epistemologists no doubt assume that in formulating the standards for epistemic rationality they are limning the rationality of belief simpliciter' (ibid.). Again, this is an analogous charge that Cocking & Kennett lay against moral theorists in arguing that the nature and good of friendship demonstrates that morality is not always overriding. However, Stroud moves a step further, so to speak, and argues that the nature of friendship calls for a re-evaluation of the standard within which we measure what we ought to believe all-things-considered. To make a similar claim in the moral domain is thus not, as Cocking & Kennett do, merely assert that morality cannot always tell us what to do all things considered, but rather to re-evaluate what it is morality asks of us.

While Stroud does not explicitly endorse this response, she does offer the thought that there could be good reasons for thinking that mainstream epistemological theories are wrong to evaluate the doxastic practices unfavourably:

For example, if we take a leaf from the moral theorist's book and shift evaluative focus points, we might see some epistemic merit in the contribution that the good friend's beliefs make to social learning and to the accuracy of the social distribution of opinion concerning her friend...A move to a social perspective could thus

highlight recognizably epistemic values that transcend the narrower confines of evidentialism ("Epistemic Partiality" 523).

As Stroud argues, such a move might not be considered from a conventional epistemic point of view, and seems to follow an 'innocent-until-proven-guilty' approach and is certainly a valuable point to make. The move to highlighting the social value of friendship in the epistemic domain – especially the 'accuracy of the social distribution of opinion' concerning one's friend – is interesting because it points to the fact that friendships are not self-contained relationships. Indeed, Stroud is not the only moral theorist to point to the social value that friendship has. Friedman (1989), although not referring specifically to the doxastic practices within friendship, argues, 'friendship can involve the mutual support of, in particular, unconventional values, which can be an important stimulus to moral progress within a community' (Helm "Friendship"). Through directing and supporting each other friends thus might come to develop a set of beliefs or moral values that are at odds with more entrenched, impersonal and abstract moral guidelines. Of course, the argument that this has social value hinges on both how these values come to affect the wider community and an assumption that moral progress is possible.²³ In other words, the social value of friendship is only valuable insofar as it has a positive (moral) effect on the community. It is difficult to see how one would argue, for example, that the relationship between two friends who together develop racist beliefs has a social value. In this case the friendship undoubtedly has individual value – as indeed it does for everyone – but it seems fair to say that the social value is lacking.

By pointing to the positive outcomes that may result from the good friend's doxastic bias Stroud argues along similar lines. She claims that because non-friends may be quick to believe the worst of your friend, your better, more authoritative opinion of him qua friend may counteract this. However, the merit of this seems to depend on 'the fact of the matter' about what your friend did. By this I mean that the social value the good friend's doxastic practices have

²¹ To illustrate this Stroud gives the following example: '[E]ven when you don't know any further specific facts...you may possess general knowledge about your friend which can help to place his actions in a different light. Suppose you know that your friend's shyness around women whom he finds attractive takes the form of a seeming standoffishness and indifference. This information allows you to entertain a different, and more accurate, interpretation of what may have seemed to the woman in question to be a haughty brush-off' (Stroud 516).

²² In order to demonstrate this position (that the appeal to a privileged perspective ultimately cannot vindicate the good friend's belief-forming process) Stroud proposes a thought-experiment in which friendship as epistemic variable is removed. In it Stroud argues that intuition tells us that we would not adopt the same doxastic practices for an ex-friend (to whom we bear no grudge) as we would for a current friend. The point here is that we possess the same amount of evidential information, but do not adopt the same doxastic practices as a consequence of this. Incidentally, one could draw two conclusions from this: that we are either being unfair to the nonfriend (and by extension to others in general) or we are being unfairly partial to our friend. My sympathies lie with the former conclusion, and indeed this is something that Stroud seems to suggest later on. For the full argument see: "Epistemic Partiality" (pp. 517-518).

²³ Whether moral progress is possible is a debate I will not entertain here. Having said this, I will make the (not uncontroversial) assumption that it is indeed possible.

seems to be contingent on the good friend's 'actual' character or what he 'actually' did. It also implies that it is possible for an agent to balance the individual value his friendship has against the social value. In other words, an appeal to the social value of the friend's doxastic practices also means they also have to be justified socially by the friend. Looking back to Stroud's earlier methodological remarks, an appeal to the social value of friendship means that the friend must be able to square his beliefs as a friend with a general consensus of what the doxastic practices should constitute. This argument seems to be evinced by the fact that Stroud argues that there is a limit to the good friend's bias. Furthermore, this point is both the one at which the friend is at which he is forced to recognize that his epistemic bias is no longer socially justifiable (as this epistemic bias necessarily involves third-parties), and the one at which the friendship relation (and its commitment) is broken. In other words, it is the point at which, as a friend, you feel not only that you should end the friendship for your own sake, but simultaneously the one at which you feel that everyone would end their friendship. This is, then, where we apply and appeal to a 'fundamental impartiality.' It is the point beyond which (if you did not break the friendship relation) you would be both immoral and you would no longer be a good friend – for example, by implicitly and unquestioningly endorsing the friend's bad actions or beliefs. This last point is particularly strong because it is through friendship that two people come to share each other's values – there is a strong sense (though, of course, not an absolute one) in which your friend's values are also your own, and seen to be so by others. By (publicly) believing something of your friend, then, you are also saying something about yourself.

DILEMMAS IN FRIENDSHIP

Perhaps it is necessary to look a couple of examples to both clarify and substantiate the dynamics of my claims here. In order to do so I now turn to two examples from film and literature respectively: John Ruane's *Death in Brunswick* (1991) and Mohsin Hamid's *Moth Smoke* (2000). By way of forewarning, I would like to point out that both the examples I use here centre on extreme conflicts between (impartial) morality and friendship. Although there are unquestionably countless other, and more mundane, reasons for someone to end a friendship other than

moral considerations, I draw attention to these more exceptional circumstances because it is not my project to provide a guide to living. Rather, my interest lies in circumstances in which people are faced with the choice to end their friendship. I use these examples because they stretch the epistemic practices of the friend to the point where such a choice becomes necessary. Such circumstances are the only way to truly illustrate a direct conflict between the individual and the social value of friendship (the social value belonging to the moral domain).

The example taken from *Death in Brunswick* is used by Cocking & Kennett in "Friendship and Moral Danger" to illustrate that friends are not always moral exemplars, and that reasons arising from friendship sometimes override those of morality as explained above. Keeping this in mind, I quote their summary in full:

Carl, the main character of the film Death in Brunswick, is no saint. Weak, vain, and disorganized, he is a severe disappointment to his mother. He drinks too much, and he works as a cook at a seedy nightclub in Brunswick where he falls foul of the owners by falling in love with a young barmaid. One night, Mustapha, his drug-dealing kitchen hand is badly beaten up in the back alley by the nightclub heavies. Carl is warned to keep his mouth shut; Mustapha is told that Carl is responsible for the beating. So late that night, Mustapha staggers into the kitchen and lunges at Carl, who is holding a long-pronged fork. Mustapha impales himself on the fork and dies. In a panic, Carl calls his best friend Dave, an easy-going family man. Against the protests of his wife, June, Dave dresses and drives to the nightclub to see what is up. His initial response when shown the body is that the police must be called. Carl begs him not to, saying that he could not cope with going to jail. Faced with Carl's fear, Dave takes charge and helps Carl move the body. They take it to the cemetery where Dave works, he breaks into a coffin in an open grave, stamps on the putrefying corpse inside to make room for Mustapha, and re-closes the coffin. Later, they deny all knowledge of Mustapha's disappearance to his distressed widow and son ("Friendship and Moral Danger" 279-280).

The tensions between (impartialist) morality and friendship that Cocking & Kennett wish to direct us to here are straightforward. It is clear that, from the perspective of morality, Dave has made a mistake.²⁴

At the same time, it seems that Dave has risen to the occasion in the way close friendship demands of us by helping Carl, an innocent man,²⁵ stay out of jail for a murder he did not commit. In fact, to put it more strongly, we might say that had Dave not done so, then we could say he had failed as a friend – and perhaps even might not be a friend to Carl at all. It is easy to see, then, why Cocking & Kennett claim that this in an example of how, and when, our commitment to our friends 'override' our commitment to morality. Before discussing the merits of this position, I now turn to the dilemma that Daru, the main protagonist of *Moth Smoke*, faces:

Darashikoh Shezad (Daru to his friends) is a university-educated junior banker living in Lahore in 1998. One day, having recently been fired for misconduct, Daru is driving around the streets of the city when he sees his wealthy and privileged best friend Ozi, whom he has known since childhood, accidentally hit a child crossing the street with his Pajero. What's more, Daru is shocked when Ozi quickly drives away from the scene of the accident. After driving the boy to hospital, Daru decides to confront Ozi at his home. When he arrives and tells Ozi that he saw what happened Ozi makes it clear that he does not particularly care about the boy, nor is he prepared to admit responsibility either to the police or the boy's family (of the latter he reassures Daru that the boy's family will be 'recompensed'). Daru cannot believe what he hears, and is overcome with fury and disbelief. He drives back to the hospital where the boy was admitted, but by the time he arrives the boy has died. When the police question him about the accident Daru tells them he does not know who hit the boy (in fact, because of this he later becomes a suspect). As for the friendship between Daru and Ozi, throughout the rest of the novel it is clear it is over and that this was the event that caused the rupture.

I chose this second example in conjunction with the one used by Cocking & Kennett as much for its differences as its similarities. The two are similar in that they both involve an accidental killing by one of the two close friends and both involve a denial

of involvement about the incident to the police and the victim's family (both by the 'perpetrator' and the helping friend). A final similarity is that, in both cases, the reader or the viewer believes (or, at least, is invited to) that, all things considered, the friend faced with the dilemma did not do the wrong thing. That is to say, if Dave has failed as a moral agent, I think it is a shared feeling that he has not failed as a friend and that, if we were in the same situation, we would at least consider acting in the same way despite any moral qualms we may have. Likewise, I think most people would agree that Daru was right to end his friendship with Ozi (perhaps even that he was required to so), not necessarily because of what Ozi *did*, but because of his indifference towards it. Whether we are (or should be) in agreement with Daru in his choice not to report Ozi to the police is a matter I will leave open. However, leaving that aside, it is the case that in both instances we are invited to accept what Dave and Daru decide, all-things-considered, to do.

This brings us to the important *differences* between *Death in Brunswick* and *Moth Smoke*, the most important one being that in the former the friendship prevails while in the latter it does not. This fact seems to turn on another difference between the two examples: Carl is genuinely upset about what happened and fears unjust repercussions, while Ozi is totally indifferent (perhaps because he knows there will be no legal consequences to his actions because of his privileged status). A third difference is that Dave does not see what happens to Carl while Daru sees everything happen before him. Lastly, there is a difference in narration between these two examples: in *Death in Brunswick* the camera does not adopt any one character's point of view and, furthermore, as Carl is the main character we see what really happens to him in the kitchen. This is important because with respect to Dave we, as the audience, are in a privileged epistemic position and so not confronted with the same burden on our beliefs. The narration in *Moth Smoke*, on the other hand, is in first-person present tense so we are limited to Daru's subjective telling of events. This means that not only can we not

²⁴ In the uncontroversial sense that secretly disposing of a corpse, defiling a grave and lying to a bereaved family constitutes a serious moral wrong.

²⁵ Innocent, at least, of premeditated murder.

move beyond Daru's perspective, we are also left with the question of whether Daru is trying to convince us that his version of events is 'actually' true. Having said this, I do not think that this need trouble us very much, and indeed it may make the discussion more fruitful. As we have access to Daru's thoughts, this merits a close look at what he thinks (or claims he does) when he confronts Ozi at his home:

The red Pajero is parked in the driveway, Ozi watching a servant wipe the dent in its bumper with a wet cloth. My best friend is wearing sunglasses, a bright T-shirt, and knee-length shorts. He looks like an overgrown child. A child who gets everything. Gets away with everything. I step out of my car as say very softly, 'I saw you just now.' For a moment he watches me, silent, expressionless, as though he's trying to remember who I am. Not remember: decide. Decide who I am (Hamid 96-97).

Now that I have described both examples let us consider what they tell us about the relation between friendship, belief and morality.

As I mentioned, Cocking & Kennett utilize the example from *Death in Brunswick* to demonstrate how our commitment to friendship can at times override our commitment to morality: Dave helps Carl despite the moral reasons he has not to. It also seems to be a good example of the epistemic bias (and its social value) that is constitutive of friendship in that Dave helps to save his friend from going to jail. However, this example is not as effective as Cocking & Kennett imagine because, in the strictest sense, Dave does not come to a different conclusion than an external observer would. To put it more precisely, there is nothing at stake for our beliefs about Carl in how Dave acts in the situation. By this I mean that, because we know the situation, and we know that Dave is a friend of Carl's, we as external observers can vouchsafe that Dave's actions are at a minimum, all things

considered, justifiable. This example does not really show that the demands of friendship *override* those of morality because Dave's doxastic bias has been taken out of the equation because we, as external observers, have the *correct* belief as we saw what happened. Dave's friendship with Carl makes his actions more coherent, but nothing more than that.

The dynamics of the dilemma in *Moth Smoke* are – for our purposes – more interesting because the epistemic bias of friendship, and the friendship itself, is broken. The passage from the novel quoted above is particularly interesting because it is the moment in which Daru is forced into the same epistemic position (or forced into the same doxastic practices) as the external observer. By this I mean that the evidence he is confronted with is sufficiently direct and overwhelming that the slanted belief-forming process Stroud describes²⁶ reaches its end-point, and is left no more wiggle-room. Daru, then, is forced, by virtue of the evidence presented to him, to draw the same conclusion as a nonfriend or external observer would. What happens as a result of Ozi's actions? First, Daru suddenly sees his friend as an overgrown, spoiled child²⁷ and then he sees that Ozi is trying to decide who Daru is to him (the implication being that this has been the case all along). In processes familiar to us from the drawing account of friendship, we see here that the moment that Ozi does something unforgivable – something that Daru cannot accept of his friend – Daru's interpretation of him changes. This is not altogether surprising. More interesting is that at this moment that Daru realizes (or claims to) that Ozi is no longer, and never has been, engaged in the mutual drawing and willingness to be interpreted that is fundamental to friendship. In fact, Cocking & Kennett argue that the lack of this is a terminating condition of friendship and whatever value the friendship had disappears. Thus, the fact that Daru (now) comes to Ozi 'pre-packaged' to conform to an image that suits

Ozi comes as the final blow in their relationship.

As I have mentioned, due to the fact that *Moth Smoke* is written in the first-person present tense we must be alive to the possibility that Daru is trying to 'convince' the implied reader of his subjective version of events. I draw attention to this because in the moment he decides the relationship is over Daru seems to go a step further than he needs to: not only does he claim that he was mistaken about Ozi, but he also implies that Ozi never was a real friend. Daru not only suddenly sees Ozi for what he 'really' is – a spoiled, oversized child – but he also perceives (or, more accurately, projects) a calculating look in Ozi that allows Daru to implicitly undermine the entire basis of their friendship. This move, then, *doubly* establishes Daru's innocence by allowing him to suggest he was not only mistaken about his friend, but also that he was manipulated into believing the friendship was genuine. In fact, this move comes across – if I may draw an analogy – as a case in which a suspect not only denies that he knows who stole the million-dollar diamond, but also that he knew the diamond was stolen at all.

This is important for several reasons (especially because of the reader's limited epistemic access in this text) and is a consequence of the idea that the differential doxastic practices have social value (as well as individual value). As I suggested a few pages above, if the differential doxastic practices that are constitutive of friendship have social value then it follows that they should be *justified* in the social domain. It is thus significant that, in the moment that he ends his friendship, Daru simultaneously justifies to himself and to the (implied) reader (who, in this case, do not have a privileged epistemic access, and so cannot 'objectively' adjudicate between competing versions of events, unlike in *Death in Brunswick*) suggests that Daru is attempting to clear his own name in the court of public opinion. Why his own name? In the drawing account we are each other's creators to some extent so Daru is trying to distance himself from Ozi to preserve his own name. Furthermore, this is exactly what ending a friendship

is: making a statement about your beliefs and values in contrast to the other, 'rejected' friend. At the same time, it also seems a last-ditch effort at bringing your friend round to his senses by showing Ozi what he will lose by going down this path (Daru 'only' ends his friendship, but does not report Ozi to the police). Daru only makes the further claim – that he and Ozi were never 'really' friends – not for himself, but for the reader, who stand in for wider society that he has internalized in his thought process. It is, then, precisely at the moment that Daru is forced to take up the same epistemic position as an external observer would that he also has to confront (or, at least is aware of) the social dimension that his friendship with Ozi had and respond to this.²⁸

What Daru's 'doubling of innocence' suggests is that friends are more reflexive about their epistemic practices than Stroud proposes. Recall that she claims that, '[a]s a good friend, your belief set is slanted: you actually believe your own spin' ("Epistemic Partiality" 513). While I agree that it is a feature – and even a duty – of friendship that you 'actually' believe the best of your friends, this does not mean that you are not aware that your beliefs are biased. To draw a parallel with the discussion about individual and social value: just as the friend can be aware of both the individual value and the social value that his friendship has, so too can the good friend honestly believe the best of his friend yet acknowledge his belief is partial. This is possible because it is not determined which of the two epistemic perspectives (i.e. that of the non-friend or of the friend) is 'objectively correct.' In other words, even if the friend is 'objectively' correct about his friend's actions (as in the case in *Death in Brunswick*) the beliefs he holds and the belief-forming practices that formed them are still biased. Therefore, it is possible for the friend to both 'really' believe his beliefs and know that they are biased (just as friend's are both able to believe in each other's moral values and still be aware that these are not widely shared). If this were not the case, then the friend's epistemic practices would be *completely* impervious to new evidence and the friend would

²⁶ The doxastic bias, as described by Stroud, of the friend is characterized by the friend's 'slowness in updating her beliefs' and a 'relative imperviousness' to new (negative) evidence without regressing to a total denial of reality.

²⁷ The child carrying connotations of being incapable of understanding moral reasoning, but also of being someone to whom morality does not apply whether he is capable of grasping it or not. This latter point is significant because it illustrates that a reason for Daru to reject Ozi is that he sees Ozi as someone who sees himself as exempt from morality.

²⁸ Friendship internally produces its own intrinsic value and systems of commitments and practices, but it is when the relation of friendship comes into contact with the social realm that this value – its 'social value' – acquires a moral flavour as it affects people outside this relation differently than those in it.

never update her belief about her friend. That is to say, if the friend had an unquestioning commitment to her friend, then the differential doxastic practices would not exist because if the evidence does not match her beliefs about her friend (and she believes that she is 'objectively' correct about her friend) she would have no reason to consider the evidence at all. It is, therefore, significant that the point at which Daru ends his friendship with Ozi he both appeals personal reasons (that he was mistaken about Ozi, and so that the personal value of the relation has been dispelled) and disavows – internally²⁹ – the entire friendship in the social realm. It is here that Daru says, metaphorically, 'I am ending my friendship now not only because Ozi violated the (moral) values we shared (or that I thought that we did), but also because this is the point at which I think everyone would (and should) end their friendship.' In a way this is an acknowledgement, or an assertion, that while the epistemic bias is internally³⁰ valuable, it is only socially justifiable as long as there is enough 'wiggle room' in the evidence about your friend to allow for the possibility that your doxastic bias is socially valuable. This in turn suggests two things: first, that while Stroud's call for a reconceptualization of 'epistemic rationality' is certainly warranted, it seems the good friend's epistemic bias is 'rational' only on the condition that that there is not enough evidence to override it. So we might instead say that, all things considered, the friend's epistemic bias is perhaps not epistemically rational, but epistemically justifiable. The friend's doxastic practices are rational only if they are done out of friendship and socially justifiable. Secondly, if friendship has social value, and that when it does it is moral, then Cocking & Kennett are wrong to claim that reasons arising from friendship can *override* those of morality.

The two examples from *Death in Brunswick* and *Moth Smoke* demonstrate that while it is true that

friends may have subjective beliefs and belief-forming processes where their friends are concerned, this does not entail that there is not a minimal epistemic standard which we must not fail to adhere to. This minimal claim comes into force when it is no longer feasible for the friend to maintain his epistemic bias. Having said this, it does not follow that, when the friend's doxastic bias is still feasible that he has most reason, all things considered, to believe the best of his friend. Instead the friend has (epistemically) sufficient reasons – i.e. reasons arising out of friendship – to hold the beliefs that he does given both the personal value and (potential) social value of these beliefs. A similar move can be made in the moral domain to reconcile moral overridingness with the nature of friendship.

CONCLUSION

By way of conclusion I would like to say a few words – and offer a few suggestions – about the conflict between moral overridingness and friendship that Cocking & Kennett point to in "Friendship and Moral Danger." Their claim is that in light of the conflict between these two evaluative perspectives (or commitments) there is no clear-cut answer to what we ought to do, all-things-considered, in a particular situation. For them the framework of morality is such that it cannot accommodate the coordinates of friendship, and so we have no reason to suppose morality is 'always' overriding. However, Cocking & Kennett's response is not the only one available to us. We may instead argue that, rather than giving up on moral overridingness, we should (as in the epistemic case) reconsider what it is that morality asks of us. Two articles that propose such a 'restructuring' of morality are Shiffrin's "Moral Overridingness and Moral Subjectivism" (1999) and Stroud's "Moral Overridingness and Moral Theory" (1998).³¹ Shiffrin rebuts a common argument that morality is not

overriding because moral reasons are subjective,³² and argues that 'a critical source of overridingness may be found in objective features of morality, at least given a reasonable conception of morality's structure' (Shiffrin 772) and so is not dependent on anyone's subjective reasons to accept it for its truth. Stroud formulates an account of moral overridingness that takes into account the sorts of objections that Cocking & Kennett raise against it, and argues why it is desirable for the thesis of moral overridingness to be sustained. This latter point seems to be demonstrated by the fact that Cocking & Kennett introduce a qualification in their rejection of moral overridingness – namely that we are not justified in committing murder for our friends. The conflict Cocking & Kennett see, rather, is that morality and friendship provide seemingly irreconcilable frameworks with which to guide our actions, and that a friend who is only willing to engage in mutual drawing to the extent allowed by morality is no true friend.

It is not difficult to be sympathetic to Cocking & Kennett's argument, and indeed it gestures toward an argument of Stroud's – that friendship involves a certain (moral) psychology: just as one would expect a friend to suspend judgement on you, so you would expect a friend to quell certain moral qualms in order to help you. However, where Stroud points to a limit in the good friend's doxastic bias (and

indeed a similar baseline in the moral realm, as does Shiffrin), Cocking & Kennett do not. They see the conflict between morality and friendship as one between two competing commitments. However, they misunderstand the difference between these two commitments. Cocking & Kennett see neither of these two commitments as *always* being stronger than the other (in our motivational set), and improperly argue that a commitment to morality (and moral overridingness) necessarily compels us to treat *moral* reasons and requirements as *all-things-considered* requirements – as reason *simpliciter*. On both counts they are wrong: in the latter case their formulation of moral overridingness is unnecessarily strong, and in the former do not do justice to the difference(s) between these two commitments. It is wrong to posit a commitment to morality (and moral overridingness) as necessarily stronger and, instead, we should see it as a deeper commitment. Our commitment to morality is rather like a ship anchored to the sea floor. The influence of our friends, personal interests and circumstances may cause us to drift across the moral plane, but the anchor ultimately prevents us from going adrift, and losing ourselves.³³ It is, therefore, problematic that Cocking & Kennett claim that our commitment to morality (or moral overridingness) leads to the loss of much of the good of friendship by subordinating our friendships to the moral

²⁹ By this I mean within Daru's own thought process, which we have access to as readers.

³⁰ By 'internally' I mean valuable to the person's involved in the friendship.

³¹ I will not go into a full-blown discussion of moral overridingness, but instead merely propose – contra Cocking & Kennett – that friendship (as described by the drawing account) is compatible with the idea that morality is overriding. The two accounts of moral overridingness that I touch upon here support this claim.

³² The claim is that 'moral considerations provide reasons for a person only if they appeal to independent aspects of her psychology, character, or life, such as her aims projects and relationships...[and so theorists who hold moral subjectivism to be true] doubt that these subjective, reason-grounding, factors will reliably be strong enough to guarantee that all-things-considered moral requirements will always override conflicting considerations' (Shiffrin 772).

³³ In a discussion about the difference between our commitment to our friends and to morality Shiffrin writes, 'One occasion, one may forsake a friend in urgent, moderate need, although friendship-generated reasons require one to give aid...But, there are limits: one cannot, consistent with one's identity as a friend, forsake the friend in urgent, moderate need to tend to a stranger in moderate need or to pursue an intriguing intellectual lead...I suggest an account of the structure of the end of morality that differs by not admitting the same degrees of freedom' (786). Shiffrin's analogy of our commitment to morality (and moral overridingness) is that of a legal contract – it is something that we must not fail to adhere to, even if we are not inclined to. In other words, once the commitment is made it is forever binding. Of course, we should only commit to such a thing only if we think its requirements are not unjust, or as Stroud puts it, it should only require a moral minimum. If Shiffrin and Stroud are right about the structure of overridingness (and also of friendship) – as I argue they are – then this way of conceptualizing the difference between the commitments complements my own analogy.

framework, and thereby doing violence to our own *identity*.³⁴ The task, then, is to formulate an account of moral overridingness that is compatible with the drawing account of friendship, and also why it is both plausible and preferable to have such an account.

In "Moral Overridingness and Moral Theory" Stroud proposes an articulation of moral overridingness that meets the first of these tasks by being sensitive to personal (circumstantial) considerations a moral agent has. Her Overridingness Thesis (OT) holds that 'moral requirements do not require us to be moral saints, but rather require a moral *minimum*: that which we morally must not fail to do' ("Moral Overridingness" 172). Stroud formulates her thesis in a way that preserves the decisive authority of morality: 'If S is morally required to \emptyset , then S has most reason to \emptyset ' (ibid.). It is important to note here that a moral requirement and what done has most reason to do are two distinct concepts. Specifically, the importance lies in the idea that a moral requirement is not a requirement *simpliciter*.³⁵ By claiming only that we have decisive reasons to do what is morally required of us, OT 'asserts only that we have compelling reasons not to do what is morally impermissible' ("Moral Overridingness" 172). Turning to what is meant by 'having the most reason to \emptyset ,' Stroud states that 'reason' here means 'reason for

action,' i.e. a consideration relevant to the generic practical question of what to do (in a particular situation)' (ibid.).³⁶ The claim of the OT, then, is that moral requirements generate decisive reasons to \emptyset within the field of reasons an agent might have to act otherwise. However, as in the epistemic case, OT does not claim that it is *irrational* to act wrongly. For example, Stroud mentions (and endorses) Michael Slote's 'satisfying' view of rationality on which it is not a rational requirement to do what one has most reason to do. On such a view of rationality one could claim it is not irrational for an agent to act wrongly despite having most reason not to – *providing the reasons are good enough*. What OT claims, then, is that within the field of reasons that act upon an agent, moral requirements provide the balance of reasons in favour of choosing a particular course of action. If, as Stroud claims, it is possible for an agent to weigh various reasons against each other, there must be some framework through which to do so. While she does not give a full explication of what such a framework might be,³⁷ she does note that it is at least plausible to believe it exists because it expresses an intelligible idea (in the sense that it is precisely these sorts of deliberations that we do in our daily lives).³⁸

Lastly, if the thesis of moral overridingness is to overcome the charges made against it by Cocking

& Kennett, it must convince us why we accept its judgement when it makes one, to convince us why Cocking & Kennett both misunderstand it (and so why they are wrong to reject it). In order to do so the OT must demonstrate both that moral requirements generate reasons for action and that, once generated, these reasons are not defeated by considerations that have personal value to the agent. If our conception of morality does not account for the non-moral reasons we have in our lives, or if it does not accord them their proper weight, then it cannot claim that morality is overriding. This is an important point because it shows that in the field of reasons that affect our actions, those that relate to our self-interest have special rational weight.³⁹ It is only when morality accords an agent's interests their *full* weight that we have reasons to commit to it and accept its adjudication and sacrifice our personal interest when necessary. This conception of morality is compatible with moral overridingness, in that it requires only a moral minimum (i.e. by not asking of us unjust sacrifices). Of course, what remains is to determine exactly how much it can ask of us, but this is not an argument that counts against the thesis that morality is overriding and so is also not defeated by the claims of its incompatibility with friendship that Cocking & Kennett raise against it.

There remains, then, the task to show why – contra Cocking & Kennett, and Cottingham – it is

desirable to have a theory of moral overridingness. In "Moral Overridingness and Moral Theory" Stroud suggests some reasons why we would benefit from holding moral overridingness. The reasons she proposes are fairly uncontroversial but nevertheless bear mentioning as it reminds us of what Cocking & Kennett claim we lose in the face of friendship. Briefly put, the reasons Stroud suggests are that morality is commonly seen as putting a constraint on our aims; secondly, 'some of us actually take moral requirements to be overriding' ("Moral Overridingness" 176) and this would seem rationally unmotivated was this not true;⁴⁰ and thirdly, it is commonly accepted that a moral requirement gives enough reason for someone to \emptyset where other evaluative perspectives may not yield sufficient reason. The danger Stroud sees in the loss of morality's overridingness is that we would lose an important (if not the most important) guide to our conduct. In a similar vein, Shiffrin also dwells on the consequences of the loss of morality's absolute (and objective) authority. She claims that our commitment to morality is the most stable basis around which to build an identity and further, 'to recognize the moral perspective as providing a comprehensive view on the weight of all relevant considerations at one moment, and then, at another, to go deaf to its dictates would risk manifesting a rather incoherent identity' (Shiffrin 792-793). This contradicts Cocking & Kennett's claim

³⁴ This line of argumentation is not unfamiliar. In "Ethics and Impartiality" (1981) and "Partiality, Favouritism and Morality" (1986) Cottingham argues that the impartiality thesis is untenable. A powerful argument he gives against (his strict interpretation of) the impartiality thesis is that 'if anyone seriously attempted to live in the way suggested, then it is doubtful whether he could survive as a person, as a whole individual, at all [my emphasis]' (ibid. 87). While Cocking & Kennett are, admittedly, not undertaking the same project as Cottingham, they share the same unease about the possibility of a valuable life as an individual within a commitment to a moral framework as necessitated by moral overridingness.

³⁵ To say that S is morally required to \emptyset is 'to come to an overall moral verdict about S's situation' (ibid.). In other words, rather than claiming an agent has moral reasons to \emptyset (say, for example, arising from some *prima facie* duty), for Stroud moral requirements already 'take into account the particular circumstances S is in now' (ibid.). This move goes some way to defeating the sorts of objections Cocking & Kennett raise with the example from *Death In Brunswick* because we can claim that Dave is not morally required to report Carl to the police (just as he was not required to believe Carl murdered Mustapha in the epistemic domain) precisely because the demands of morality are alive to these sorts of circumstances and the weight of personal, non-moral reasons for action. ³⁶ That is to say, 'reason' here is (approximately) an all-things-considered reason, not a consideration that is relevant from the point of view of a particular domain, or system of evaluation (e.g. morality or justice).

³⁶ That is to say, 'reason' here is (approximately) an all-things-considered reason, not a consideration that is relevant from the point of view of a particular domain, or system of evaluation (e.g. morality or justice).

³⁷ However, for a full investigation of precisely such a framework that is sympathetic to Stroud's claims, see: McLeod, Owen. "Just Plain 'Ought.'" *The Journal of Ethics* 5.4 (2001): 269-291.

³⁸ A plausible account of this was given in chapter 2 of this thesis, where we see how (and why) Daru balances his commitment to his friendship with Ozi and his commitment to his moral values. In this chapter I suggested that Daru's (and the good friend's in general) doxastic bias – and specifically its limit – would not exist if the good friend is not aware of, and responsive to, external beliefs about his friend. Similarly, to make sense of the social value (and the entailing justification) of friendship there must be a conception of moral overridingness that is absolute. That is to say, if (impartial) morality is not always overriding, Daru's 'doubling of innocence' is incoherent because there would be no reason for him to do so.

³⁹ Shiffrin proposes the same argument in order to make the authority of morality plausible. See p. 792 of "Moral Overridingness and Moral Subjectivism."

that a friend who would only engage in the mutual drawing and interpretation characteristic of close friendship to the extent that this was permissible within the moral framework is no true friend (and thus suffers the adverse effects of this with regard to her identity or character development). However, this claim is based on an incorrect interpretation of what it is that morality asks of us. Once this is corrected, not only does this argument lose its purchase, but we see the converse is true – that not to engage in mutual drawing and interpretation to the extent this is allowed by the parameters limned by OT would be to lose much of the good of friendship.

To see this I return a last time to the moment that the friendship between Daru and Ozi ends. As I argued, in this moment Daru is forced to draw the same conclusion about Ozi as would an external observer and – as a consequence of this – claims he was mistaken both about his friend and their friendship (his ‘doubling of innocence’). Specifically, the way Daru describes the ‘realization’ that his friendship with Ozi was never a true friendship is that he ‘sees’ Ozi deciding who he is (i.e. that Ozi is not really engaged in their friendship as prescribed by the drawing account). This shows that, in the moment that Ozi has undeniably failed to do what is morally required of him, Daru perceives this as a threat to his autonomy (over his own identity) and reacts accordingly. This reaction is to assert his autonomy against Ozi’s influence through an appeal to the social (moral) domain of friendship (i.e. that his commitment to, or friendship with, Ozi is no longer socially justifiable). This reaction makes sense given a little introspection of what ending a close friendship is. Ending a friendship is an amoral act that is in a strong sense demanded by morality: Daru ends his friendship to preserve his (moral) identity, but does not report Ozi to the police (for which he arguably has moral reasons to do). In accepting Ozi despite the grave moral wrongs he has committed – to go ‘deaf to the dictates of morality’ here, as Shiffrin puts it, despite the concessions to the personal reasons we have it makes – Daru would truly lose himself. Thus, it seems that we lose much of the good of friendship *not* if we conduct them *within* the parameters of

morality, but rather outside them. A few days after the confrontation between Daru and Ozi, Ozi comes to Daru’s house to check whether he told the police about the accident. When Daru tells him he hasn’t Ozi thanks him and adds, “I must admit, I’ve been pissed off with you. I didn’t like the way you acted. It wasn’t what I expected from a friend.” [...] “We’re not the boys we were when we were seventeen,” Ozi says, “But my view on friendship hasn’t changed. Friends support each other no matter what. Do you agree?” (Hamid 140). Daru concludes, as should we, that he does not because this is against the very nature of friendship.

WORKS CITED

Aristotle. *The Nicomachean Ethics*. Trans. Robert. C Bartlett and Susan D. Collins. Chicago: University of Chicago Press, 2011. Print.

Baron, Marcia. “Impartiality and Friendship.” *Ethics* 101.4 (1991): 836-857. Web.

Bernstein, Mark. “Friends Without Favouritism.” *The Journal of Value Inquiry* 41 (2007): 59-76. Web.

Cocking, Dean and Jeanette Kennet. “Friendship and the Self.” *Ethics* 108.3 (1998): 502-507. Web.

Cocking, Dean and Jeanette Kennett. “Friendship and Moral Danger.” *Journal of Philosophy* 97 (2000): 278-296. Web

Cottingham, John. “Ethics and Impartiality.” *Philosophical Studies* 43.1 (1981): 83-99. Web.

Cottingham, John. “Partiality, Favouritism and Morality.” *The Philosophical Quarterly* 36.144 (1986): 357-373. Web.

Death In Brunswick. Dir. John Ruane. Perf. Sam Neil, Zoe Carides and John Clarke. Australian Film Finance Corporation, 1991. Film.

Friedman, M.A. “Friendship and Moral Growth.” *Journal of Value Inquiry* 23 (1989): 3-13. Web.

Grunebaum, James. “Fair-Weather Friendships.” *The Journal of Value Inquiry* 39 (2005): 203-214. Web.

Hamid, Mohsin. *Moth Smoke*. New York: Picador, 2000. Print.

Helm, Bennett. “Friendship.” *The Stanford Encyclopaedia of Philosophy* (2009) Ed. Edward N. Zalta. Stanford University. Web. 21 Feb. 2012.
<http://plato.stanford.edu/archives/fall2009/entries/friendship/>

Jeske, Diane. “Friendship, Virtue, and Impartiality.” *Philosophy and Phenomenological Research* 57.1 (1997): 51-72. Web.

McLeod, Owen. “Just Plain ‘Ought’.” *The Journal of Ethics* 5.4 (2001): 269-291. Web.

Shiffrin, Seana. “Moral Overridingness and Moral Subjectivism.” *Ethics* 109.4 (1999): 772-794. Web.

Stocker, Michael. “The Schizophrenia of Modern Ethical Theories.” *The Journal of Philosophy* 73.14 (1976): 453-466. Web.

Stocker, Michael. “Values and Purposes: The Limits of Teleology and the Ends of Friendship.” *The Journal of Philosophy* 78 (1981): 747-65. Web.

Stroud, Sarah. “Epistemic Partiality in Friendship.” *Ethics* 116.3 (2006): 498-524. Web.

Stroud, Sarah. “Moral Overridingness and Moral Theory.” *Pacific Philosophical Quarterly* 79 (1998): 170-189. Web.

Vice, Samantha. “Truth and Love Together at Last: Style, Form, and Moral Vision in ‘Age of Iron’.” *J.M Coetzee and Ethics, Philosophical Perspectives on Literature*. Eds Anton Leist & Peter Singer. New York: Columbia University Press, 2010. 293-317. Print.

Wolf, Susan. “Morality and Partiality.” *Philosophical Perspectives* 6 (1992): 243-259. Web.

⁴⁰ This argument follows a similar line to why Daru’s ‘doubling of innocence’ would be incoherent were moral overridingness not to be true.

The Effects of Serotonin in the Nucleus Accumbens Shell on Glucose Metabolism in Rats

Kayleigh van Megen



ABSTRACT

Rationale: Recently, a patient with type 2 diabetes mellitus (T2DM) was treated with deep brain stimulation (DBS) in the nucleus accumbens. After DBS treatment, the glucose metabolism of the patient was altered. To study the mechanism by which DBS influences glucose metabolism, this technique was applied in an animal model. DBS of the nucleus accumbens shell (AcbSh) in rats rapidly increased blood glucose and plasma glucagon concentrations in a region- and intensity dependent manner. Previously, it was observed that the neurotransmitter serotonin is increased in response to DBS. In the current study, we hypothesized that the neurobiological mechanism underlying this AcbSh-DBS effect on glucose metabolism involves modulation of serotonin levels.

Objective: Our aim is to study the effects of serotonin in the AcbSh on glucose metabolism.

Methods: Male Wistar rats received bilateral microdialysis probes in the AcbSh in addition to a jugular vein and carotid artery catheter. We subjected the rats to 60 minutes of reverse microdialysis of either Ringer solution (vehicle) or the selective serotonin re-uptake inhibitor fluoxetine (83 μ M or 250 μ M). Blood samples were drawn before, during, and after cessation of drug administration to measure blood glucose, and plasma insulin concentrations. Endogenous glucose production was measured by stable isotope dilution.

Results: Reverse microdialysis of fluoxetine significantly increased blood glucose levels as compared to vehicle infusion. The other parameters measured were not altered.

Conclusion: These preliminary data support a role for serotonin in the AcbSh-DBS effects on glucose metabolism. Future research should focus on the specificity of action of serotonin to the AcbSh, and the brain structures it signals to in order to affect glucose metabolism.

1. INTRODUCTION

We are our brain: that was the mindset of the 1990s, called the Decade of the Brain. Neurologists believed that almost every aspect of human functioning could be explained by a phenomenon in the brain (Jones and Mendell, 1999). With the turn of the century, a rather disturbing trend caught the attention of us all. In the 2000s the debate about the increased prevalence of

obesity and diabetes became unstoppable. With the arrival of a new decade came a much-anticipated interdisciplinary approach to medical problems, and by now the scientific world has accumulated strong evidence for a role of the brain in the development and progression of diabetes (Sandoval et al., 2009).

Recently, a patient with type 2 diabetes mellitus (T2DM) was treated with deep brain stimulation

(DBS) in the nucleus accumbens. DBS is used as a treatment for psychiatric disorders like major depressive disorder (van Dijk et al., 2011), but its mechanism of action is yet unclear (Sesia et al., 2010). After DBS treatment, glucose metabolism of the patient was altered (Case report, Metabolic Unit AMC, unpublished data). The nucleus accumbens is known to regulate behavior, especially reward. Dysfunction of the reward system is known to play a role in obesity (Wang et al., 2001), which in turn is associated with the development of T2DM (Ye et al., 2011). Whether the nucleus accumbens has an independent role in the disturbances in glucose metabolism observed in obesity is unknown.

Several studies have shown that there is an anatomical route, which might underlie part of the effects of DBS on glucose metabolism. Viral tracing experiments revealed a neural connection between the nucleus accumbens and the pancreas (Buijs et al., 2001). Furthermore, Stratford and Kelley (1999) found that the shell region of the nucleus accumbens (AcbSh) has neural projections to the lateral hypothalamus (LH), a brain structure involved in glucose metabolism (Wade et al., 2008). Whether this neuro-anatomical route is functional in glucose metabolism, was studied by use of an animal model. In this study, rats were subjected to acute DBS in the AcbSh at two different intensities (100 and 200 μ A) (Diepenbroek et al., unpublished data). In addition, each rat was sham treated, i.e. the same animal was subjected to the DBS ON- and, at a later time point, to the DBS OFF procedure (sham), and therefore functioned as its own control. The animals stimulated with 200 μ A showed significantly increased levels of glucose, glucagon, corticosterone, but no effect on insulin, compared to their sham situation. The rats stimulated with 100 μ A had increased levels of corticosterone, but no difference in the other parameters, compared to their sham control. Since in both intensity groups corticosterone levels were raised, it is unlikely that corticosterone causes the increase in glucose levels. Furthermore, rats with misplaced electrodes did not have increased glucose levels, while corticosterone was increased. Thus, the animal model showed that bilateral DBS in the AcbSh at the highest stimulation intensity increased blood glucose concentrations and plasma glucagon concentrations.

Now that a link between DBS and glucose

metabolism was verified, the question remained what mechanism, triggered by DBS, caused the rise in concentrations of blood glucose and plasma glucagon. DBS in the AcbSh has been shown to increase levels of serotonin in the rat brain (Sesia et al., 2010). In addition, researchers have proposed a monoamine deficiency hypothesis for the development of major depressive disorder, which is treated with DBS (Belmaker and Agam, 2008), indicating that DBS might increase serotonin levels. Interestingly, previous research has proven that serotonin is involved in energy balance and appetite (Lam et al., 2010). It is yet unclear, however, if serotonin is the neurotransmitter responsible for the change in glucose metabolism in response to DBS.

Our research aims to delineate this gap in knowledge, by answering the following research question: Does modulation of serotonin levels in the AcbSh have an effect on glucose metabolism? We aimed to answer this question by use of the selective serotonin reuptake inhibitor (SSRI) fluoxetine. Fluoxetine increases extracellular serotonin levels dose-dependently when administered to the rat brain (Rutter and Auerbach, 1993). Fluoxetine was administered in the AcbSh of rats via reverse microdialysis. Prior, during and after administration, blood glucose, plasma concentrations of the glucoregulatory hormone insulin, and endogenous glucose production (EGP) were measured.

In order to position this research in the field, a literature review is given prior to the description of the current research. After this, the materials and methods of this study will be followed by the results. An in-depth discussion will elucidate the implications of these results. Future research will be proposed in the conclusion, which recapitulates the data presented.

LITERATURE REVIEW

1. PREVALENCE OF TYPE 2 DIABETES MELLITUS

Type 2 diabetes mellitus (T2DM) has become a global problem. Although predominantly seen as a Western societal issue, in 2010 6.4 % of the world population was suffering from T2DM (Ye et al., 2011). As this prevalence is reaching remarkable numbers, it can hardly be said that T2DM is a problem of Western societies alone. Diabetes prevalence in China is rising tremendously, and other non-Western countries are following (Yang et al., 2010). It is hard to envision the seriousness of the situation, since diagnosis is rather poor in undeveloped countries (Aekplakorn et al., 2003). Besides being a clinical problem, T2DM is also costly. In 2010, related spendings amounted to \$376 billion (Ye et al., 2011). The brain has been implied in the regulation of glucose metabolism and is proposed to play a role in the development and progression of diabetes (Sandoval et al., 2009). The precise mechanism underlying such a regulation is still largely unknown.

2. DEEP BRAIN STIMULATION AND GLUCOSE METABOLISM

The deep brain stimulation (DBS) animal study (Diepenbroek et al., unpublished data), described in the Introduction, supports a role of the brain in the control of glucose metabolism. DBS is nowadays widely applied as a treatment for psychiatric disorders like obsessive compulsive disorder and major depressive disorder (van Dijk et al., 2011), but its mechanism of action is unclear (Sesia et al., 2010).

DBS has different effects when applied to different brain areas (Kringelbach et al., 2007). Van Dijk et al. (2011), for instance, concluded that DBS in the nucleus accumbens core does not affect monoamine release, whereas Sesia et al. (2010) found that DBS in the nucleus accumbens shell (AcbSh) does increase dopamine and serotonin levels. Therefore, it can be hypothesized that changes in neurotransmitter levels released from the AcbSh as a consequence of DBS could be involved in the effects on glucose metabolism. Furthermore, Stratford et al. (1999) blocked γ -aminobutyric acid (GABA) fibers from the AcbSh projecting to the medial ventral pallidum. The neurons from the medial ventral pallidum terminate in the lateral hypothalamus (LH), which is known to be involved in the control of food intake (Wade et al., 2008). This blockade induced increased

food intake in already satiated rats (Stratford et al., 1999). Thus, not only serotonin and dopamine are released from the AcbSh, but the neurotransmitter GABA as well and its release has an effect on food intake. In addition, Buijs et al. (2001) found that the AcbSh has a neural connection with the pancreas, suggesting an anatomical route that could explain the effect of DBS in the AcbSh on glucose metabolism.

To sum up, these studies showed various alterations in neurotransmitters as a consequence of DBS. Both serotonin and dopamine levels are increased in response to DBS in the AcbSh and it was proven that there are also GABA- containing fibers in the AcbSh. Since DBS is given as a therapy for major depressive disorder, and serotonin levels are decreased in this condition, DBS probably has a profound effect on serotonin (McIntyre et al., 2006). Furthermore, serotonin has previously been proven to be involved in food intake and body weight (Lam et al., 2010). Therefore, we considered it relevant to investigate the role of serotonin in the AcbSh on glucose metabolism.

3. SEROTONIN AND GLUCOSE METABOLISM

In 1999, Horacek and his colleagues were one of the first to report a significant correlation between central serotonin activity and peripheral insulin sensitivity. No correlation was found between serotonin activity and either body weight, body mass index, or waist and hip circumference. This means that weight was not responsible for the observed response to serotonin, and that there is a direct link between serotonin and insulin sensitivity.

In more recent studies, fluoxetine, a selective serotonin re-uptake inhibitor (SSRI), was assessed as a possible therapy for type 2 diabetes mellitus (T2DM). Originally, fluoxetine, more commonly known under the tradename Prozac, was used as an antidepressant (McIntyre et al., 2006). Ye et al. (2011) reviewed several studies in which T2DM patients without diagnosis of depression were treated with fluoxetine. Overall, fluoxetine caused modest but significant weight loss and significant decreases of fasting plasma glucose and glycated hemoglobin (HbA1c), a measure of average blood glucose levels over several months prior to sampling. Fluoxetine was well tolerated and there were no major adverse effects. As Horacek et al., Ye and colleagues supported the hypothesis that the observed metabolic benefit is not solely due to

a decrease in body weight, and suggested a role for insulin as a mediator of this metabolic effect.

McIntyre et al. (2006) reviewed several antidepressants, among them serotonergic antidepressants like fluoxetine, with regard to their metabolic effects. Fluoxetine treatment resulted in a significant decrease in HbA1c and lowered insulin requirements in T2DM patients. Overall, they concluded that serotonergic antidepressants increase insulin sensitivity, decrease blood glucose levels, and do not have an effect on body weight.

The above reviewed experiments indicate a role for fluoxetine in the improvement of glucose metabolism. Unfortunately, the site of action of fluoxetine in these experiments is unclear. Fluoxetine is given orally and therefore it could have its effect in various sites in the body. In addition to central serotonin activity, there is evidence for peripheral serotonin activity, which is supposed to be beneficial for glucose metabolism as well (Watanabe et al., 2011), and it is not possible to draw conclusions about central serotonin activity from these results. Nevertheless, fluoxetine has been shown to cross the blood-brain barrier (Henry et al., 2005), and have a central effect (Rutter and Auerbach, 1993), thus contributing to its use as an antidepressant.

4. THE SEROTONIN 2C RECEPTOR AND GLUCOSE METABOLISM

In order to find out how serotonin causes an effect on glucose metabolism, Bonasera and Tecott (2000) studied serotonin receptor function. Up to date, 14 distinct subtypes of the serotonin receptor have been identified (Wade et al., 2008). By looking at serotonin receptor mutants, Bonasera and Tecott (2000) found that the serotonin 2C receptor (5HT-2CR) is involved in the negative regulation of feeding and that 5HT-2CR knockout mice display a T2DM syndrome. 5HT-2CR is expressed both in and outside of the hypothalamus, but is restricted to the central nervous system (Wade et al., 2007). More specifically, Eberle-Wang et al. (1997) located the 5HT-2CR in the nucleus accumbens. Interestingly, Zhou et al. (2007) showed that 5HT-2CR agonists administered to a mice model of T2DM improved their glucose tolerance. In addition, the improvement in glucose metabolism was dependent upon the activation of melanocortin-4 receptors, which are important in maintaining insulin sensitivity (Heijboer et al., 2005). In accordance with

this, Xu et al. (2008) conducted an experiment in which they globally knocked out the 5HT-2CR, which resulted in hyperphagia and obesity. They found that the re-expression of 5HT-2CRs solely in pro-opiomelanocortin (POMC) neurons, which have previously been implied in the regulation of appetite (Lam et al., 2010), was sufficient to normalize the hyperphagia and obesity. In a very recent publication, Papazoglou et al. (2012) demonstrated that POMC hypothalamic neurons express 5HT-2CRs as well as insulin receptors and the authors hypothesized a cross-talk between these two signaling pathways. This is supported by studies by Orosco and colleagues (2000; 2001), who showed that in response to a carbohydrate meal serotonin is released from the hypothalamus, which causes an increase in hypothalamic insulin levels. In Papazoglou's experiment, a diabetic rat model had reduced hypothalamic serotonin release in response to a meal. This suggests that the onset of T2DM may be accompanied by changes in the brain (Papazoglou et al., 2012).

5. THE LINK BETWEEN TYPE 2 DIABETES MELLITUS AND MAJOR DEPRESSIVE DISORDER

Interestingly, the 5HT-2CR is not only implicated in T2DM, but also in major depressive disorder (MDD) (Dracheva et al., 2008). Iwamoto and colleagues (2005) studied the degree of RNA editing of the 5HT-2CR in depressed rats. RNA editing of the 5HT-2CR results in different isoforms of the receptor; the higher the degree of RNA editing, the lower the functionality of the receptor (Lyddon et al., 2011). Using the learned helplessness model of depression, Iwamoto concluded that depressed rats show a higher degree of RNA editing as compared to control rats. Furthermore, fluoxetine treatment of the depressed rats decreased the degree of RNA editing of the receptor (Iwamoto et al., 2005). A post-mortem human study by Lyddon and colleagues (2011) found that RNA editing is only increased in MDD patients who committed suicide, and not in MDD patients who died from other causes. Thus, the association between the 5HT-2CR and depression remains controversial.

Nevertheless, a link between MDD and T2DM has been proposed before. Silva et al. (2012) reviewed the association between depression and insulin resistance. The data was inconsistent. They conclude with a positive note, however, since clinical trials assessing the effect of depression treatment on

insulin resistance produce positive results (Silva et al., 2012). Two other recent studies, a case-control and a cohort study, independently concluded that the use of fluoxetine as an antidepressant significantly increased the risk of developing T2DM (Khoza et al., 2012; Pan et al., 2012). Unfortunately, these studies merely discuss correlation, but not causation. This correlation could imply that the use of fluoxetine leads to an increased risk of developing T2DM. It could be, however, that fluoxetine treatment and T2DM are related through a third factor, i.e. the neural substrates shared by MDD and T2DM, e.g. the 5HT-2CR. Thus, if this were the case, the neural overlap between the two diseases would predispose one to developing both conditions.

6. THE GAP IN KNOWLEDGE

The literature reviewed above clearly implies a key role of serotonin in the regulation of glucose metabolism. Whether serotonin is the main player involved in altering glucose metabolism in response to DBS remains unknown. Although much ground has been covered on the effects of serotonin on glucose metabolism, the mechanisms underlying these effects still have to be clarified. It is, for instance, still unclear from which precise brain structure serotonin signals for glucose metabolism. The current study hypothesizes that serotonin affects glucose metabolism, as is seen in response to DBS, via the AcSh.

MATERIALS AND METHODS

1. ANIMALS

Thirty-five male Wistar rats (250-280 g) (Harlan, Horst, the Netherlands) were individually housed in Plexiglas cages in a temperature (20 ± 2°C), humidity (60 ± 2%), and light controlled room with a 12/12h light-dark schedule (lights on at 7:00h). Half of the rats were subjected to fluoxetine microdialysis infusion in one of two doses (250 and 83 µM), while the other half received a vehicle infusion (Ringer) solution. When not subjected to the microdialysis, the rats were unrestrained and both chow (SDS, England), and tap water were available ad libitum. The experiment was approved by the Committee for Animal Experimentation of the Academic Medical Centre of the University of Amsterdam, the Netherlands.

2. SURGERY

While anaesthetized with an i.p. injection of 80 mg/kg ketamine (Eurovet Animal Health, Bladel, the

Netherlands), 8 mg/kg Rompun® (Bayer Health Care, Mijdrecht, the Netherlands) and 0.1 mg/kg atropine (Pharmachemie B.V., Haarlem, the Netherlands), intra-atrial silicone catheters were inserted in the vena jugularis and carotid artery of the rats. In addition, microdialysis probes were bilaterally inserted, aimed at the AcSh (A +1.44 mm, L +3 mm, V -7.3 mm, angle 17°), using a stereotaxic apparatus (Kopf). Catheters and microdialysis probes were fixed on to the skull with dental cement. Prior and the day after surgery, the animals received a 0.3 ml subcutaneous injection of the non-steroidal anti-inflammatory drug Rimadyl® (Pfizer Animal Health B.V., Capelle aan de IJssel, the Netherlands) (20x diluted in saline). The rats were allowed to recover for seven days.

3. REVERSE MICRODIALYSIS OF FLUOXETINE

After the recovery period, the animals were connected to a multi-channel fluid infusion swivel. At 8:00 am the following morning, the animals were subjected to food restrictions, after which catheters were connected and continuous infusion of Ringer solution (3 µmol/h) through the microdialysis probes was started. To study glucose kinetics, the isotope [6,6-2H₂] glucose (as a primed bolus (3000 µmol/h in 5 min) infusion, followed by a continuous (500 µmol/h) infusion throughout the experiment) was used as a tracer. Before the start of the isotope infusion, a blood sample (175-200 µl) was taken at 10:00 AM to correct for background (BG) isotopic enrichment (See: Analytic methods). Three basal blood samples (B1, B2, B3) were taken after 90 minutes (175-200 µL) to measure basal glucose levels.

Subsequently, basal endogenous glucose production (EGP), and basal insulin levels were measured from the isolated plasma of the blood samples.

Subsequently, at 1:00 pm (t=0) the microdialysis infusion with fluoxetine (either 83 µM or 250 µM) started in the treatment groups and blood samples were taken at time points t=5, 10, 15, 20, 30, and 60 min (175-200 µL). Concentrations of fluoxetine were chosen on basis of previous research using in vivo fluoxetine reverse microdialysis in rats (Taylor et al., 2004). Ringer solution infusion continued in the control group throughout the experiment. At the same time points blood samples were taken from the control group. At 2:00 PM, fluoxetine infusion was switched to Ringer in the treatment groups, after which three extra blood

samples were taken at time points t=65, 90, and 120 min, in order to determine normalization of glucose and insulin levels. The blood of the control group was sampled at the same time points.

4. ANALYTICAL METHODS

A custom glucose meter was used to measure blood glucose levels (Freestyle Freedom Lite, Abbot, Hoofddorp, the Netherlands). Blood samples were immediately chilled on ice in Eppendorf tubes with 5 μ L heparin: saline (10x) solution and centrifuged at 4°C (15 min, 3000 rpm). Plasma was isolated from the blood samples and stored at -20°C until further analysis of insulin and endogenous glucose production. Plasma insulin concentrations were measured using a radioimmunoassay kit (Millipore, St Charles, MO, USA and Biochemicals, Costa Mesa, CA). The prescribed amounts of sample, standards, label, antibody and precipitating reagent of the manufacturer's protocol were divided by four, since the amount of plasma of each sample was limited. Plasma [6,6-2H2] glucose enrichment was measured by gas chromatography-mass spectrometry (GCMS) [Ackermans ea 2001]. Endogenous glucose production was calculated using Steele equations (Steele, 1959).

5. PROBE VERIFICATION

After the experiment, the animals received a single dose of pentobarbital (100-150 mg/kg BW), after which they were decapitated. Brains were dissected, frozen and stored at -80°C. To verify microdialysis probe placement, the brains of the rats were serially cut on a cryostat into 35 μ m coronal sections at -20 °C. The nucleus accumbens and the lateral hypothalamus were both collected on slides, which were stored at -80 °C. Following this, the sections were stained with thionine (see Appendix A), a metachromatic stain, and analyzed under a microscope (Leica). Since the probe is inserted into the brain, it will leave a trace in the brain tissue upon removal. Therefore, the area where the probe used to be will not be stained. A rat brain atlas (Paxinos and Watson, 1998) was used during microscopic analysis to verify probe placement.

6. DATA ANALYSIS

The data for glucose, insulin, and endogenous glucose production of each animal were analyzed with regard to their treatment situation (fluoxetine-treated vs. control), and probe placement. Verification of probe

placement will indicate whether fluoxetine-induced effects on glucose metabolism are specific for bilateral AcbSh microdialysis.

Since we have multiple measurements over time per animal, a repeated-measure analysis of variance (rmANOVA) was used to examine the effects of time, infusion and time*infusion interaction. The criterion for significance was set at $p < 0.05$. Statistical analysis was performed using SPSS (SPSS Inc, Chicago, USA).

RESULTS

1. ANIMALS, SURGERY AND PROBE VERIFICATION

Out of the 35 male Wistar rats, four animals died during surgery. Three animals were excluded from the experiment due to clogged catheters. Overall, 28 rats were subjected to the experiment. In three animals (n=1, Ringer; n=1, 83 μ M; n=1, 250 μ M), microdialysis probe placement could not be verified due to absence of traces in the brain tissue. Of the remaining 25 rats, probe verification revealed correct bilaterally placed probes in eight animals (n=2, Ringer; n=2, 83 μ M; n=4, 250 μ M). Thirteen animals (n=6, Ringer; n= 2, 83 μ M; n=5, 250 μ M) had unilaterally AcbSh placed probes. Four animals (n= 2, Ringer; n=2, 250 μ M) had bilaterally placed probes outside the AcbSh. Measurements of insulin are absent for one animal due to defrosted plasma.

Since the results of the measurements of the experiment seemed similar between the bilaterally and unilaterally placed groups, these groups were combined in data analysis. The data from the rats with misplaced microdialysis probes were not included in the analysis. Table 1 shows the final number of rats per group used for analysis.

	Glucose (mmol/L)	Insulin (ng/mL)	EGP (μ mol/kg.min)
Vehicle	8	7	8
83 μ M	4	4	4
250 μ M	9	8	9
Total number of rats	21	19	21

Table 1: Total number of rats per treatment group and parameters measured included in the analysis.

2. CONCENTRATIONS OF BLOOD GLUCOSE, PLASMA INSULIN AND ENDOGENOUS GLUCOSE PRODUCTION

In the 83 μ M infusion group, neither time nor treatment showed a significant effect on blood glucose levels. Blood glucose levels significantly increased during 250 μ M fluoxetine infusion, compared to vehicle infusion ($p=0.05$). Statistical analysis showed a significant effect of time in the 250 μ M infusion group ($p=0.002$) (Fig. 1).

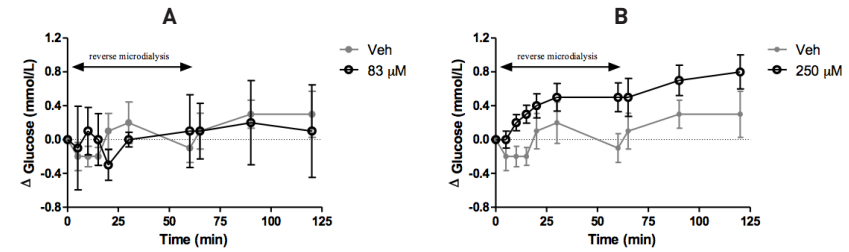


Figure 1: Blood glucose levels during and after reverse fluoxetine microdialysis in (A) the 83 μ M group (n=4) compared to vehicle infusion (n=8), and (B) the 250 μ M group (n=9) compared to the same vehicle group.

Plasma insulin concentrations showed no significant difference with 83 μ M or 250 μ M fluoxetine infusion from the vehicle infusion (Fig.2).

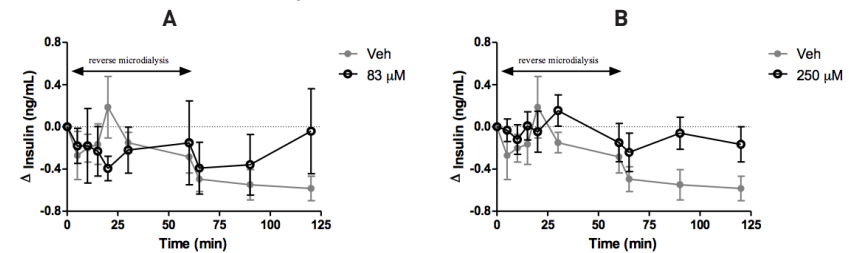


Figure 2: Plasma insulin levels during and after reverse fluoxetine microdialysis in (A) the 83 μ M group (n=4) compared to vehicle infusion (n=7), and (B) the 250 μ M group (n=8) compared to the same vehicle group.

Although endogenous glucose production (EGP) significantly decreased over time following both 83 μ M and 250 μ M fluoxetine infusion ($p=0.028$; $p < 0.001$, respectively), there were no significant differences in EGP values compared to the vehicle infusion group (Fig.3).

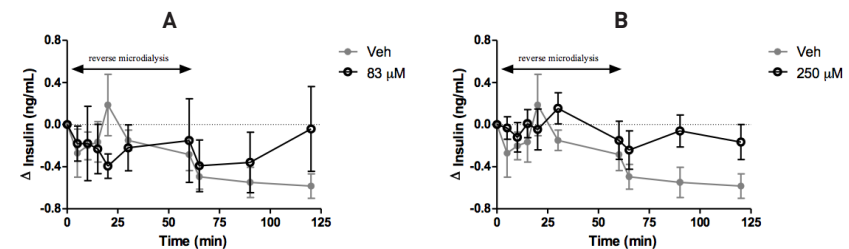


Figure 3: Endogenous glucose production [EGP] during and after reverse fluoxetine microdialysis in (A) the 83 μ M group (n=4) compared to vehicle infusion (n=8), and (B) the 250 μ M group (n=9) compared to the same vehicle group.

DISCUSSION

1. INTERPRETATION AND FUTURE RESEARCH

In this study, we proposed that AcbSh-DBS induced effects on glucose metabolism might be mediated via a DBS-induced modulation of serotonin levels, and investigated the role of this neurotransmitter in the AcbSh on blood glucose levels, plasma insulin concentrations and endogenous glucose production (EGP). We showed that reverse microdialysis of the SSRI fluoxetine significantly increased blood glucose levels.

The most prominent effect of reverse fluoxetine microdialysis on glucose metabolism was on blood glucose levels. Blood glucose levels significantly increased in the 250 μ M group as compared to the vehicle group, whereas insulin levels and EGP did not change. This is a remarkable result since the increased levels of glucose could be explained by a decrease in insulin or an increase in EGP, but neither is the case. There seems to be a trend of increased EGP in the 250 μ M group as compared to the vehicle situation, but this does not reach statistical significance. Thus, the question remains what could be the cause of the observed increase in glucose levels.

In line with the DBS animal study (Diepenbroek et al., unpublished data), as mentioned in the Introduction, we observed an increase in glucose levels. Next to insulin, other glucoregulatory hormones were measured in the DBS study, like glucagon and corticosterone. In that study, glucagon levels were increased, which could very well explain the increase in glucose levels. Therefore, in the future, also glucagon levels will be analyzed in the current study. The potential influence of stress, as measured by corticosterone levels, on glucose levels will also be further analyzed (Ulrich-Lai and Herman, 2009). Since corticosterone was not responsible for the rise in glucose levels in the DBS experiment, we hypothesize that this is also not the case in the current study.

Our data show that the fluoxetine concentration of 83 μ M was too low to reach statistical significance. This infusion concentration was included as an experiment to see whether 1/3 of the normal infusion concentration (250 μ M) would be sufficient to affect glucose metabolism. We conclude that only the 250 μ M fluoxetine infusion group provides significant information about the effect of serotonin on glucose metabolism.

In conclusion, this study has shown that reverse fluoxetine microdialysis in the AcbSh, by

increasing extracellular levels of serotonin (Rutter and Auerbach, 1993), increases blood glucose levels. The similarities between the current experiment and the DBS experiment suggest a prominent role for serotonin in the effect of DBS in the AcbSh on glucose metabolism. Nevertheless, other neurotransmitters could function synergistically with serotonin to cause a greater effect on glucose metabolism. Sandoval et al. (2009) hypothesize a cross-talk between the serotonergic and dopaminergic pathways in mediating glucose metabolism. For instance, dopamine-containing neurons in the hypothalamus are innervated by pro-opio-melanocortin (POMC) neurons (Sandoval et al., 2009), which contain serotonin 2C receptors (5HT-2cRs) (Papazoglou et al., 2012). In addition, Wang et al. (2001) found, by making use of positron emission tomography, that overall dopamine D2 receptor availability in the brain was decreased in obese individuals. Thus, a decreased expression of both receptors, 5HT-2cRs and D2 receptors, could potentially delay the transmission of the interlinked pathways with detrimental consequences. GABA signaling in the AcbSh has also been implicated in glucose metabolism, especially in the hedonic aspects of food (Volkow et al., 2011). Since the role of GABA in glucose metabolism is not much studied yet, it could be an interesting future research direction.

2. IMPLICATIONS

Our results imply that serotonin in the AcbSh affects glucose metabolism. This effect could be mediated by a neural connection of the AcbSh with the pancreas (Buijs et al., 2001), possibly via hypothalamic POMC neurons with 5HT-2CRs (Papazoglou et al., 2012). As yet, we did not conduct c-Fos staining, a technique that detects neuronal activity, in different brain areas to determine to which brain structures serotonin in the AcbSh signals for glucose metabolism. Thus, further research is needed to elucidate this topic.

Overall, our results have implications in various fields. The finding that serotonin signals via the AcbSh for glucose metabolism supports the hypothesis of a role of the reward system in glucose metabolism. Indeed, Volkow and Wise (2005) suggested the existence of a polygenic genotype that renders persons at risk for both obesity and addiction. Hayes and colleagues (2009) found that, in rats, 5HT-2CR agonists inhibit ventral tegmental area intracranial self-administration, which is a

measure of reward-related behavior. It could be that in obesity the 5HT-2CRs are less functional (Xu et al., 2008; Bonasera and Tecott, 2000) and therefore there is less inhibition of reward-related behavior, like eating high-fat and high-carbohydrate foods. In line with this, Cunningham and colleagues (2011) found that a selective 5HT-2CR agonist dose-dependently decreased the rewarding capacities of cocaine, as well as sucrose. Unexpectedly, however, Hayes et al. (2009) observed that 5HT-2CR agonists specifically administered to the AcbSh did not induce a significant effect on reward-related behavior. Although serotonin has been implicated in obsessions and compulsions (Nutt, 2008), serotonin in the AcbSh is possibly not directly involved in regulating reward, but more research should be conducted to elucidate this topic.

Interestingly, 5HT-2CR agonists have recently been considered as a novel anti-obesity treatment (Garfield and Heisler, 2009). Since obesity is one of the most reliable risk factors of developing type 2 diabetes mellitus (T2DM), fighting obesity is of utmost importance. The economic burden of T2DM is expected to rise from \$376 billion in 2010 to \$490 billion by 2030 (Ye et al., 2011). Thus, prevention of T2DM is essential, especially in the current climate of economic crisis. By avoiding obesity, as much as 80% of diabetes cases could be prevented (Ye et al., 2011). This underlines the importance of gaining a better understanding of the role of serotonin in glucose metabolism and obesity.

Our results might also give clues about the link between T2DM and major depressive disorder (MDD). We have shown that serotonin is involved in glucose metabolism, and it was previously reported that MDD patients have decreased levels of serotonin (Belmaker and Agam, 2008). Both disorders have a strong genetic and environmental component (Kahn et al., 2006; Aan het Rot et al., 2009). It could be that T2DM and MDD have common neural substrates, like decreased levels of 5HT-2CRs that predispose one for both disorders. In addition, T2DM and MDD share environmental risk factors, like stress (Heraclides et al., 2009; Aan het Rot et al., 2009). In fact, 20% of T2DM patients are diagnosed with MDD (Ye et al., 2011). In a cohort study of 2460 T2DM patients it was observed that once patients are depressed, the depression is chronic (Nefs et al., 2012). Health professionals should be aware of the link between T2DM and MDD, since treatment of co-morbid

depression does not only improve depressive symptoms but also glycaemic control (Stoop et al., 2011). Furthermore, in an effort towards personalized medicine (Hamburg and Collins, 2010), health professionals should differentiate between medication offered to T2DM patients with MDD, and patients with solely MDD. The antidepressant fluoxetine, or Prozac, is the preferential drug for the former group, since it also has beneficial effects on glucose metabolism, whereas other antidepressants do not (Ye et al., 2011).

3. LIMITATIONS

Initially, the probe verification was also intended to provide information about the specificity of serotonin action to the AcbSh. It could be, for instance, that serotonin induces similar effects on glucose metabolism when infused in other brain areas than the AcbSh. Unfortunately, this experiment failed to draw conclusions on specificity, since our misplaced probe group was too small (n=4).

The most prominent limitation to this research is of translational concern, however (Mankoff et al., 2004). Clinical studies in the same field forged rather different results than ours. Nevertheless, there is a consensus about the hypothesis that serotonin plays a role in glucose metabolism. Overall, the clinical studies define this effect as increasing insulin sensitivity and decreasing glucose levels (Horacek et al., 1999; Ye et al., 2011; McIntyre et al., 2006). Our animal study, on the other hand, found that fluoxetine in the AcbSh increases glucose levels. Except from the obvious translational concern, the different experimental observations could be explained by three discrepancies between our study and the other studies.

Firstly, the condition of our subjects, namely 'normal' Wistar rats, is different from the other studies, which mainly included diabetic (human) subjects. Secondly, the rats in our study and the DBS study by Diepenbroek et al. (unpublished data) were acutely treated by reverse microdialysis or DBS, respectively, whereas other studies implemented chronic treatment. So far, we did not measure corticosterone levels to determine if acute treatment resulted in a stressful situation for the rats, which could have an influence on glucose metabolism. Thirdly, as mentioned previously, the clinical studies administered fluoxetine orally, whereas our study infused fluoxetine specifically to the AcbSh.

Although our results were not in line with the clinical studies, this does not diminish the significance of our study. This study set out to determine the effects of serotonin specifically in the AcbSh on glucose metabolism and indeed an effect was observed. This result will contribute to the development of a general model of glucose metabolism that includes the brain amongst the other glucose-regulating organs.

CONCLUSION

Reverse fluoxetine microdialysis in the AcbSh of rats significantly increases blood glucose levels. This effect could not be explained by the other parameters measured, namely plasma insulin levels or endogenous glucose production. Nonetheless, based on the data of the DBS experiment by Diepenbroek et al. (unpublished data), we hypothesize that a rise in glucagon may be responsible for the increased levels of glucose. Therefore, glucagon measurements will be the next step in the current research.

To conclude, this study supports the link between glucose metabolism and serotonin in the AcbSh. In order to fine-tune this link, some future experiments are recommended. Currently, we

observed that acute bi- and unilateral administration of the SSRI fluoxetine in the AcbSh increases blood glucose levels, but to which brain structures it signals to induce this change remains unknown. An attempt to answer this question could come from c-Fos staining of brain structures like the lateral hypothalamus (LH), because it is hypothesized that the AcbSh projects to the LH. A double stain could be used to identify specific neurons, like pro-opio-melanocortin neurons, that are activated in these brain regions. In addition, specificity of the action of serotonin to the AcbSh should be investigated. Since our misplaced probe placement group is small, this is an option for future research of other research groups.

This study has presented only a tiny piece of the puzzle of the brain involved in glucose metabolism. Although we are a bit further in our understanding of the role of the brain, the extent of its role remains to be elucidated. As discussed, current research in this field has shown much inconsistent information. As long as we do not reach a consensus model of the role of the brain, it will be very hard to use our knowledge to invent new treatments. Therefore, research into the basic neural mechanisms underlying glucose metabolism is of utmost importance.

REFERENCES

- Aan het Rot M, Mathew SJ, Charney DS. Neurobiological mechanisms in major depressive disorder. *Canadian Medical Association Journal* 2009 ; 180(3): 305-313.
- Aekplakorn W, Stolk RP, Neal B, Wongpaisal PS, Chongsuvivatwong V, Cheepudomwit S, Woodward M. The Prevalence and Management of Diabetes in Thai Adults. *Diabetes Care* 2003; 26; 2758–2763.
- Belmaker RH, Agam G. Major Depressive Disorder. *New England Journal of Medicine* 2008;358 (1): 47-60.
- Bonasera SJ, Tecott LH. Mouse models of serotonin receptor function: Towards a genetic dissection of serotonin systems. *Pharmacology & Therapeutics* 2000; 88 (2); 133-142.
- Buijs RM, Chun SJ, Nijima A, Romijn HJ, Nagai K. Parasympathetic and Sympathetic Control of the Pancreas: A Role for the Suprachiasmatic Nucleus and Other Hypothalamic Centers That Are Involved in the Regulation of Food Intake. *The Journal of Comparative Neurology* 2001; 431: 405-423.
- Cunningham KA, Fox RG, Anastasio NC, Bubar MJ, Stutz SJ, Moeller FG, Gilbertson SR, Rosenzweig-Lipson S. Selective serotonin 5-HT 2C receptor activation suppresses the reinforcing efficacy of cocaine and sucrose but differentially affects the incentive salience value of cocaine- vs. sucrose-associated cues. *Neuropharmacology* 2011; 61:513-523.
- Dracheva S, Patel N, Woo DA, Marcus SM, Siever LJ, Haroutunian V. Increased serotonin 2C receptor mRNA editing: a possible risk factor for suicide. *Molecular Psychiatry* 2008; 13: 1001–1010.
- Eberle-Wang K, Mikeladze Z, Uryu K, Chesselet MF. Pattern of Expression of the Serotonin 2C Receptor Messenger RNA in the Basal Ganglia of Adult Rats. *The Journal of Comparative Neurology* 1997; 384:233–247.
- Garfield AS, Heisler LK. Pharmacological targeting of the serotonergic system for the treatment of obesity. *Journal of Physiology* 2009; 587 (1): 49–60.
- Hamburg, MA, Collins, FS. The Path to Personalized Medicine. *New England Journal of Medicine* 2010; DOI: 10.1056/nejmp1006304.
- Hayes DJ, Clements R, Greenshaw AJ. Effects of systemic and intra-nucleus accumbens 5HT2C receptor compounds on ventral tegmental area self-stimulation thresholds in rats. *Psychopharmacology* 2009; 203:579–588.
- Heijboer AC, Van den Hoek AM, Pijl H, Voshol PJ, Havekes LM, Romijn JM, Corssmit EPM. Intracerebroventricular administration of melanotan II increases insulin sensitivity of glucose disposal in mice. *Diabetologia* 2005; 48: 1621-1626.
- Henry ME, Schmidt ME, Hennen J, Villafuerte RA, Butman ML, Tran P, Kerner LT, Cohen B, Renshaw PF. A Comparison of Brain and Serum Pharmacokinetics of R-Fluoxetine and Racemic Fluoxetine: A 19-F MRS Study. *Neuropsychopharmacology* 2005; 30:1576–1583.
- Heraclides A, Chandola T, Witte DR, Brunner EJ. Psychosocial Stress at Work Doubles the Risk of Type 2 Diabetes in Middle-Aged Women. *Diabetes Care* 2009; 32(12): 2230-2235.
- Horacek J, Kuzmiakova M, Hoschl C, Ande M, Bahbon R. The relationship between centralserotonergic activity and insulin sensitivity in healthy volunteers. *Psychoneuroendocrinology* 1999; 24: 785–797.
- Iwamoto K, Nakatani N, Bundo M, Yoshikawa T, Kato T. Altered RNA editing of serotonin 2C receptor in a rat model of depression. *Neuroscience Research* 2005; 53: 69–76.
- Jones EG, Mendell, LM. Assessing the Decade of the Science 1999; 284 (5415): 739.

- Kahn SE, Hull RL, Utzschneider KM, Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* 2006; 444: 840-846.
- Khoza S, Barner JC, Bohman TM, Rascati K, Lawson K, Wilson JP. Use of antidepressants and the risk of type 2 diabetes mellitus: a nested case-control study. *International Journal of Clinical Pharmacology* 2012; DOI 10.1007/s11096-012-9608-4.
- Kringelbach ML, Jenkinson N, Owen SLF, Aziz TZ. Translational principles of deep brain stimulation. *Nature Review Neuroscience* 2007 (published online); doi:10.1038/nrn2196.
- Lam DD, Garfield AS, Marston OJ, Shaw J, Heisler LK. Brain serotonin system in the coordination of food intake and body weight. *Pharmacology, Biochemistry and Behavior* 2010; 97: 84-91.
- Lyddon R, Dwork AJ, Keddache M, Siever LJ, Dracheva S. Serotonin 2c receptor RNA editing in major depression and suicide. *The World Journal of Biological Psychiatry* 2011; Early Only, 1-12.
- Mankoff SP, Brander C, Ferrone S, Marincola FM. Lost in Translation: Obstacles to Translational Medicine. *Journal of Translational Medicine* 2004; 2: 14.
- McIntyre RS, Soczynska JK, Konarski JZ, Kennedy SH. The effect of antidepressants on glucose homeostasis and insulin sensitivity: synthesis and mechanisms. *Expert Opinion Drug Safety* 2006; 5(1):157-168.
- Nefs G, Pouwer F, Denollet J, Pop V. The course of depressive symptoms in primary care patients with type 2 diabetes: results from the Diabetes, Depression, Type D Personality Zuidoost-Brabant (DiaDDZoB) Study. *Diabetologia* 2012; 55:608-616.
- Nutt DJ. Relationship of neurotransmitters to symptoms of Major Depressive Disorder. *Journal of Clinical Psychiatry* 2008; 69 [Supplement E1]: 4-7.
- Orosco M, Rouch C, Gerozissis K. Activation of hypothalamic insulin by serotonin is the primary event of the insulin-serotonin interaction involved in the control of feeding. *Brain Research* 2000; 872 (1-2): 64-70.
- Orosco M, Gerozissis K. Macronutrient-induced cascade of events leading to parallel change in hypothalamic serotonin and insulin. *Neuroscience and Biobehavioral Reviews* 2001; 25 (2): 167-174.
- Pan A, Sun Q, Okereke OJ., Rexrode KM, Rubin RR, Lucas M, Willett WC, Manson JE, Hu FB. Use of antidepressant medication and risk of type 2 diabetes: results from three cohorts of US adults. *Diabetologia* 2012; 55:63-72.
- Papazoglou I, Berthou F, Vicaire N, Rouch C, Markaki EM, Bailbe D, Portha B, Taouis M, Gerozissis K. Hypothalamic serotonin-insulin signaling cross-talk and alterations in a type 2 diabetic model. *Molecular and Cellular Endocrinology* 2012; 350 (1): 136-144.
- Paxinos G, Watson C. (1998). A stereotaxic atlas of the rat brain. New York: Academic.
- Poortvliet MC, Schrijvers CTM, Baan CA. Diabetes in Nederland: Omvang, risicofactoren en gevolgen, nu en in de toekomst. Rijksinstituut voor Volksgezondheid en Milieu (RIVM) 2007.
- Rutter JJ, Auerbach SJ. Acute Uptake inhibition Increases Extracellular Serotonin in the Rat Forebrain. *The Journal of Pharmacology and Experimental Therapeutics* 1993; 265 (3): 1319-1324.
- Sandoval DA, Obici S, Seeley RJ. Targeting the CNS to treat type 2 diabetes. *Nature* 2009; 8:386-398.
- Sesia T, Bulthuis V, Tan S, Lim LW, Vlamings R, Blokland A, Steinbusch HWM, Sharp T, Visser-Vandewalle V, Temel Y. Deep brain stimulation of the nucleus accumbens shell increases impulsive behavior and tissue levels of dopamine and serotonin. *Experimental Neurology* 2010; 225: 302-309.
- Silva N, Atlantis E, Ismail K. A Review of the Association Between Depression and Insulin Resistance: Pitfalls of Secondary Analyses or a Promising New Approach to Prevention of Type 2 Diabetes? *Current Psychiatry Reports* 2012; 14:8-14.
- Steele R. Influences of Glucose Loading and of Injected Insulin on Hepatic Glucose Output. *Annals New York Academy of Sciences* 1959: 420-430.
- Stoop CH, Spek VRM, Pop VJM, Pouwer F. Disease management for co-morbid depression and anxiety in diabetes mellitus: design of a randomised controlled trial in primary care. *BMC Family Practice* 2011; 12:139.
- Stratford TR, Kelley AE. Evidence of a Functional Relationship between the Nucleus Accumbens Shell and Lateral Hypothalamus Subserving the Control of Feeding Behavior. *The Journal of Neuroscience* 1999; 24:11040-11048.
- Stratford TR, Kelley AE, Simansky KJ. Blockade of GABA receptors in the medial ventral pallidum elicits feeding in satiated rats. *Brain Research* 1999; 825: 199-203.
- Taylor GT, Farr S, Klinga K, Weiss J. Chronic fluoxetine suppresses circulating estrogen and the enhanced spatial learning of estrogen-treated ovariectomized rats. *Psychoneuroendocrinology* 2004; 29 (10): 1241-1249.
- Ulrich Lai YM, Herman JP. Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience* 2009; 10: 397-409.
- Van Dijk A, Mason O, Klompmaker AA, Feenstra MGP, Denys D. Unilateral deep brain stimulation in the nucleus accumbens core does not affect local monoamine release. *Journal of Neuroscience Methods* 2011; 202: 113-118.
- Volkow ND, Wise RA. How can drug addiction help us understand obesity? *Nature Neuroscience* 2005; 8 (5): 555-560.
- Volkow ND, Wang GJ, Baler RD. Reward, dopamine and the control of food intake: implications for obesity. *Trends in Cognitive Sciences* 2011; 15 (1): 37-46.
- Wade JM, Juneja P, MacKay AW, Graham J, Havel PJ, Tecott LH, Goulding EH. Synergistic Impairment of Glucose Homeostasis in ob/ob Mice Lacking Functional Serotonin 2C Receptors. *Endocrinology* 2008; 149(3): 955-961.
- Wang G, Volkow ND, Logan J, Pappas NR, Wong CT, Zhu W, Netusil N, Fowler JS. Brain dopamine and obesity. *Lancet* 2001; 357: 354-57.
- Watanabe H, Rose MT, Aso H. Role of peripheral serotonin in glucose and lipid metabolism. *Current Opinion in Lipidology* 2011; 22:186-191.
- Wen L, Ley RE, Volchkov PV, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Susan Wong F, Szot GL, Bluestone JA, Gordon JI, Chervonsky AF. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* 2008; 455(7216): 1109-1113.
- Xu Y, Jones JE, Kohno D, Williams KW, Lee CE, Choi Mj, Anderson JG, Heisler LK, Zigman JM, Lowell BB, Elmquist JK. 5-HT2CRs Expressed by Pro-Opiomelanocortin Neurons Regulate Energy Homeostasis. *Neuron* 2008; 60(4-2): 582-589.
- Yang W, Lu J, Weng J, Jia, W, Ji, L, Xiao J, Shan Z, Liu J, Tian H, Ji, Q, Zhu D, Ge, J, Lin L, Chen, L, Guo M, Zhao Z, Li Q, Zhou Z, Shan G, He J. Prevalence of Diabetes among Men and Women in China. *The New England Journal of Medicine* 2010; 362:1090-1101.
- Ye Z, Chen L, Yang Z, Li Q, Huang Y, He M, Zhang S, Zhang Z, Wang X, Zhao W, Hu J, Liu C, Qu S, Hu R. Metabolic Effects of Fluoxetine in Adults with Type 2 Diabetes Mellitus: A Meta-Analysis of Randomized Placebo-Controlled Trials. *PLoS ONE* 2011; 6 (7): e21551.
- Zhou L, Sutton GM, Rochford JJ, Semple RK, Lam RR, Oksanen LJ, Thornton-Jones ZD, Clifton PG, Yueh C, Evans ML, McCrimmon RJ, Elmquist JK, Butler AA, Heisler LK. Serotonin 2C Receptor Agonists Improve Type 2 Diabetes via Melanocortin-4 Receptor Signaling Pathways. *Cell Metabolism* 2007; 6(5): 398-405.

