Observing and influencing preferences in real time

Gaze, morality and dynamic decision-making

Philip Pärnamets

LUND UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Humanities, Lund University, Sweden.
To be defended at LUX, room C121, Lund, April 18, 2015, at 13:15.

Faculty opponent

Alan Kingstone

Lund University Cognitive Science
Department of Philosophy
LUND UNIVERSITY

Document name

Doctoral dissertation

Date of issue 2015-04-18

Author(s) Philip Pärnamets

Sponsoring organization

Title and subtitle: Observing and manipulating preferences in real time: Gaze, morality and dynamic decision-making

Abstract: Preference formation and choice are dynamic cognitive processes arising from interactions between decision-makers and their immediate choice environment. This thesis examines how preferences and decisions are played out in visual attention, captured by eye-movements, as well as in group contexts.

Papers I-II make use of the Choice Blindness paradigm. Paper I compares participants' eye movements and pupil dilation over the course of a trial when participants detect and fail to detect the false feedback concerning their choices. Results indicate objective markers of detection with important implications for questions concerning possible demand effect or cognitive dissonance explanations of choice blindness. Paper II examines another aspect of the choice environment, namely, the social context. Choice blindness is demonstrated in small groups for the first time. It is shown that preferential change can be induced in dyads by manipulating the group's beliefs about their choices, thus extending the preference change through choice effects beyond individuals for the first time.

Paper III examines how visual attention differentially supports both decision and memory processes depending on the amount of task-relevant information available to participants. Participants' performance and visual attention dynamics was found to vary depending on the amount of information available to them, and the results indicate that decision outcomes are heavily influenced by encoding prior to traditional choice phases used in decision research.

Papers IV-VII concern decision-making in the moral domain. Paper IV investigates visual attention when participants choose between difficult moral dilemmas, showing asymmetries in how participants distribute their attention depending on making utilitarian or deontological choices. Paper V introduces a novel paradigm for influencing decision based manipulating the timing of decision by measuring the direction of gaze while the participant deliberates. Using this method participants' moral decisions were biased without their knowledge to a randomly chosen alternative. This shows that moral decision and visual attention are highly coupled and that by simply knowing where someone is looking it is possible to influence their decision process. Papers VI&VII build on the links between gaze direction and moral choice and present the first computational decision models based on eye gaze applicable to the moral domain. Paper VI for choices between abstract principles and Paper VII for donations to charitable organizations.

Together the papers advance novel methodological solutions to understanding preferences and decisions across a number of domains, both highlighting the important contributions of social and sensorimotor interactions to the content of our decisions as they develop over time, as well as demonstrating how decisions can be influenced by leveraging those interactions.

Key words: eye gaze, morality, decision-making, preference formation

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

ISSN and key title: 1101-8453 Lund University Cognitive Studies 160

ISBN: 978-91-7623-241-5

Recipient's notes

Number of pages 270

Price

Security classification

Signature

Date 2015-03-04

Observing and influencing preferences in real time

Gaze, morality and dynamic decision-making

Philip Pärnamets

LUND UNIVERSITY

Cover art by Josef Mellergård.

Cognitive Science
Department of Philosophy
Faculties of Humanities and Theology

ISBN 978-91-7623-241-5 (print)
978-91-7623-242-2 (pdf)
ISSN 1101-8453
Lund University Cognitive Studies 160

Printed in Sweden by Media-Tryck, Lund University
Lund 2015

Acknowledgments

One of the great privileges of being a PhD student, apart from being allowed to work long, dark hours in one's heart's delight to discover something about something probably very few others care about, is all the people you encounter during that time. Not only have I had the joy of meeting fantastically brilliant people from whom I have learnt a great deal, but I have had the chance to make new friends, but also to rediscover my existing ones – so often forgotten in lieu of yet another tantalising deadline or project – and to find my family always there where I momentarily had left them. This thesis is to all of you since without you none of this would have happened anyway.

I have had the joy of having been under the tender care of three brilliant mentors and supervisor.

Lars Hall, you have taught me far more than you imagine, from your striking way with words to always asking me just the question I least wanted to hear at the moment I probably least expected it. You have always pushed me the extra mile in my thinking and for that I am beyond grateful.

Christian Balkenius, you have been a stable haven throughout my years, always providing me a sense of calm. I value your vision and clarity, and our discussions have always made me strive harder to match your insights.

Petter Johansson, I don't know what to say – without your friendship, support and trust I would not have come anywhere at all. Domo arigato, sensei! I hope to make you proud. (Here I should also thank Marie and Moltas for letting me borrow Petter so many times the past months!)

I spent a significant part of 2012 at University College London, which proved to be one of the most stimulating times of my years as a PhD. Most importantly, I had the benefit to get to know and work with Daniel Richardson. I am humbled and thankful for have been allowed this opportunity to work with you Daniel, it has been truly fantastic (PS I still have your lab keys!).

While at UCL, I was lucky enough to work with Jorina von Zimmermann and Ramsey Raafat, who introduced me to the wonderful world of groups. In this context, I also had the joy to benefit from comments and thoughts from Nick Chater, who saved our work many times.

Much of my time at UCL was spent in and around room 201 (with special thanks to Dave Lagnado allowing me space!), where I met a great deal of fantastic people: Tobias Gerstenberg, Christos Bechlivanidis, Neil Bramley, Paula Parpart, Costi Rezlescu, Keith O'Brien, Lukasz Kopec and Adam Harris, as well as, Matt Gobel and Chris Street. Special mention to Marsha Kirichek for letting me to test some wild ideas and be part of your thesis work. Thanks to all!

I have also had the great fortune to get to know Michael Spivey. I have learned much from you and our conversations have fuelled much of my perspective on the human mind.

Much thanks to Ulrike Hahn and Rick Dale, and a wine-fuelled conversation with both of you which helped channel much of my thoughts about the meaning of my main paper.

The bulk of my PhD years have been spent here in Lund at LUCS. Here is my second home. I cannot begin to explain what a wonderful and stimulating environment it is, and what a fantastic workplace; thanks to everyone past and present: Agneta Gulz, Björn Sjödén, Paulina Lindström, Magnus Haake, Åsa Harvard, Richard Andersson, Magnus Johnsson, Mathias Osvath, Helena Osvath, Jens Nirme, Kristin Ingvarsdottir, Zahra Gharaee, Can Kabadayi, Ivo Jacobs, Tomas Persson, Elainie Madsen, Jana Holsanova, Peter Gärdenfors, Betty Tärning, Joel Parthemore and Birger Johansson. Special thanks to Annika Wallin, for always going the extra mile to engage me in exciting new projects and to teach me lessons from the real world outside the lab. Kerstin Gidlöf and Roger Johansson, for a great collaboration. Rasmus Bååth for always being willing to discuss (and help with!) the nerdier sides of statistics.

The department would grind to a halt without some people hard at work keeping it afloat: Agneta Ahlberg, Anna Cagnan Enhörning, Eva Sjöstrand, Anna Östberg, Björn Petersson and Tomas Persson. Many, many thanks for help with all things large and small!

A lot of my data recording has been done in and around the Humanities Lab and using the eye-tracking equipment residing there. There I found an excellent research environment and the ever so excellent eye-tracking seminars. Kenneth

Holmqvist deserves as special thanks. I am also much indebted to Marcus Nyström for always being willing to help at critical times, and the constant encouragement you have given me over the years! Likewise, without the advice of Richard Andersson starting as a master student in lab, I would not have come far! And of course all the participants in the seminars over the years for valuable feedback and discussions.

I want to thank Emmanuel Genot and Justine Jacot for being such wonderful creatures. Organising conferences, endlessly redrafting papers and eating canard washed down with bourbon has never been more fun!

With Thomas Strandberg and Andreas Lind I have not only been sharing the same lab, supervisors and frustrations, but also many joys in and outside of work. Thomas, together was have gorged in fantastic films and eaten ourselves halfway to a mortal sin – I can't imagine a better person to do this with! Andreas, your constant encouragement, warnings that things are not always what they seem and the fantastic turns of our discussions have been invaluable!

My friends outside of work have given me much support, and living together with many of them in Rymdblomman has been a source of strength over and over again: Jonathan Mellergård, Kaj Svedberg (rabu rabu), Oskar Hallström, Gustav Nilsson, Sofia Nordstrand, Sara Andersson and Victor Rydberg – woop woop! In addition, a special thanks to Oliver Vikbladh for sharing your mind with me, Noomi Weinberg for many fun travels and late nights, Andreas Holmgren for always bringing my ego back to earth, Johanna Landberg for always being there, Josef Mellergård for letting me decompress so many times in your home, and David Ivarsson for all our adventures in space.

I also wish to thank Pekka Mellergård for reminding me about perspective and your friendly questions.

The dancefloor has been my therapy room, spa and funhouse and many nights and early mornings have been spent in attics, warehouses and cellars. To the people behind and in front of the decks – one love.

My family has been fully supportive over the years; despite my sometime inabilities to actual explain what I am doing and why it takes so much time. To my mom and dad especially – thank you for letting me become me.

Finally, the last years have been much brighter since I met Filippa – being with you means the world to me.

Contents

List of original papers

Paper I

Pärnamets, P., Hall, L., Strandberg, T., Balkenius, C., & Johansson, P. (submitted). Looking at choice blindness: Evidence from eye-tracking and pupil dilation.

Paper II

Pärnamets, P., von Zimmermann, J., Hall, L., Raafat, R., Chater, N., & Johansson, P. (submitted). Choice-induced preference change in groups.

Paper III

Pärnamets, P., Johansson, R., Gidlöf, K. & Wallin, A. (submitted). How information availability interacts with visual attention during judgment and decision tasks.

Paper IV

Pärnamets, P., Hall, L., & Johansson, P. (submitted). I see your dilemma: Visual attention and moral choice.

Paper V

Pärnamets, P., Johansson, P., Balkenius, C., Hall, L., Spivey, M.J. & Richardson, D.C. (in press). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences.*

Paper VI

Pärnamets, P., Balkenius, C. & Richardson, D. C. (2014). Modelling moral choice as a diffusion process dependent on visual fixations. In Bello, P., Guarini, M., McShane, M. & Scassellati, B. (eds.) *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Austin, TX.

Paper VII

Pärnamets, P. (2015) A fixation dependent decision model of charitable choice. *Lund University Cognitive Studies 161*.

Scope and summary

Imagine walking down a street with a friend, casually chatting as you make your way towards a movie theatre. Just outside the theatre you are stopped by a homeless person asking for spare change. While stopping and considering you look back and forth between the person and your goal, the movie theatre. Suddenly, your friend tugs your arm and hurries you to move on, less you will miss the beginning of the film. At that moment you are forced to make your decision, to give the homeless person change or not. There might be all sorts of reasons going through your head at that moment for doing one thing or the other. Some of these might have influenced your final decision, or at least you could argue so. However, here I want to ask a different question: could the precise moment when your friend tugged your arm have affected your choice? And, if so, did the direction you were looking at when you were interrupt have any influence?

The answer to both these questions, as you will discover is 'Yes'[1]. More generally, in this thesis several investigations into the interactions between social and sensorimotor processes, on the one hand, and preferences and choices, on the other, are presented. The work grew out of a seemingly innocuous question; can something (useful) be learnt about moral choices by studying eye movements? What particularly interested me was how eye movements might reflect ongoing deliberation and choice. Over the years, while working on this thesis, the question grew and mutated, seemingly on its own - like an untended lunch box in the fridge - into a broader interest concerning how preferences and choices evolve in general. Eventually I found myself equipped with a dynamic perspective on human cognition and having done a number of thematically and methodologically linked studies.

[1] See Paper V

The example above was chosen not only because it illustrates the main findings of what I consider to be the flagship paper of this thesis (Paper V), but also because it captures the main research themes of the papers and the introductory chapters. These themes can be summed up as dynamic cognition, decisions, and morality.

First, throughout I assume a perspective of cognition as being continuous and embodied. The key lesson from such an approach is that it emphasis the time course of cognitive processes, how they develop over time and how this can inform our understanding of the mind. In particular, in most papers this takes the form of studying eye gaze and how it dynamically relates to choice and preference formation in various ways.

Second, all the studies herein concern decisions and preferences. In Papers II and V this takes the form of studying how preferences can be influenced. In the former, by manipulating a group's beliefs about its own choices, while in the latter, by using knowledge about where someone is looking to interrupt and bias their decision process. Paper I investigates how participants choices concerning false feedback about previous decisions evolve. In the remaining papers choices are studied as they unfold in real-time by investigating the time course of eye-movements in various ways.

Third, this thesis addresses questions concerning moral cognition. It does so by treating the moral deliberation of the agent as a decision. This might not seem as a radical stance, but surprisingly few studies in moral psychology have examined how choices between alternatives are made in contrast to how participants make judgments about the appropriateness of some scenario or other. In the four papers concerning moral cognition, I have studied moral choices between options. The aim has been to investigate parts of our complex moral psychologies by treating moral choices as choices like any others, and see how they might be revealed by eye gaze.

Over the introductory chapters, these three themes will be explored in greater depth, setting the stage for the papers, but also coherently developing the implications that I draw from the results. In particular, I believe that the papers in this thesis inform our understanding of human cognition in three different ways. First, they demonstrate the viability of approaching cognition from a dynamic perspective generally, and explore new areas to do this in. Second, the findings here support a constructed preference view of decision-making, and provide novel reasons for adopting that view. Third, they demonstrate that eye gaze can be used to productively study moral cognition and show that moral choices are in part constituted by gaze dependent decision mechanisms.

Before proceeding I will below summarise the papers included and give a short terminological explanation.

A brief note on terminology

As many of the papers here concern research that has involved eye-tracking a clarifying note on terminology is in order. Throughout the thesis I will refer to participants looking at things. To do so, most often the term 'gaze' or 'eye gaze' is used. Other times, in particular in Paper III, 'visual attention' is used. Sometimes, but more rarely 'fixation' is used, most often in connexion with the modelling work. Fixations are oculomotor events, and as such are used as inputs when considering participants' decision processes in the models discussed. When it matters for the argument to speak of fixations, that term is used, but otherwise, especially in this introduction, most of the discussion will concern the direction participants are looking in, which is taken to be where they also are attending. The relationship between attention and eye gaze is discussed in greater detail in the next chapter.

Summary of papers

This thesis consists of seven papers, and the main results and contributions of each paper are summarised here to aid the reader's understanding of the introductory chapters.

The papers are organised thematically but not chronologically. First are three papers on non-moral choices, of which the first two also concern the choice blindness methodology. The papers can be read as they are presented or in any order. They are followed by four papers concerning moral choices. These, especially Papers V-VII are more closely linked, and might benefit from a sequential reading.

Paper I - Looking at choice blindness: Evidence from eye-tracking and pupil dilation.

Research question

Choice blindness is the finding that people are sometimes blind to mismatches between intended choices and their outcomes – meaning that we can be experimentally manipulated to accept false feedback about our choices and led to willingly confabulate reasons about choices we, in fact, never made. In this paper, we investigated participants' eye movements and pupil dilation during the presentation of false feedback in a choice blindness task. We wanted to see if differences could be found between detected and non-detected manipulated trials which would give objective evidence concerning claims that participants really are choice blind in choice blindness studies.

Procedure

Seventy-six participants were recruited to take part in a facial preference study. Participants sat in front of a computer and made choices between pairwise presented faces. Their eye-movements were monitored by a tower mounted eye-tracker. Following each choice the chosen face was presented to them a second time and they were asked to indicate which one of six presented facial features had contributed the most to their choice. Participants could also choose a seventh option; that they preferred the other face. On eight out of thirty-six trials, participants were given false feedback about their choice so that their originally non-chosen face was presented to them as their chosen one.

Results and Conclusion

Overall participants detected 59% of the manipulations. Analysis showed that during detected trials participants responded faster, that their pupils were significantly more dilated compared to when failing to detect. Additionally participants looked less at the face and more at the response options when detecting compared to when not. Together this shows how processing differences characterise detections in a choice blindness task, which helps rule out demand effect and cognitive dissonance explanations of the choice blindness paradigm.

Paper II - Choice-induced preference change in groups

Research question

Choices are known to affect preferences in individuals. In this paper, we investigated if group choices would affect later preferences of the group. To do this, we utilised the choice blindness methodology to induce false beliefs about choices made by a dyad. As a corollary, we also wanted to determine if choice blindness could be found for group choices at all, and compared detection rates to that of individuals in identical tasks.

Procedure

In the first experiment we studied thirty-six dyads and forty individuals in a simple choice blindness experiment. Participants were told that they would choose between pairs of faces, which person they thought would make the best flatmate. Following their choice, participants' chosen face was presented a second time and participants were told to discuss (dyads) or state (individuals) why they chose that person. Participants were given false feedback about their choices on four out of twenty trials.

In the second experiment forty dyads were studied in the same experiment as above but with one important addition. Following the first phase of twenty choices (four manipulated) the dyads were asked in a new, second phase to choose again for eight of the face pairs they had originally chosen. The eight second round trials included the four manipulated pairs from the first phase.

Results and Conclusion

In the first experiment we found that dyads would accept the false feedback and accept the manipulated choices as their own. However, dyads did exhibit higher detection rates compared to individuals (35% compared to 20%).

In the second experiment, we found that dyads changed their preference in the second phase as a result of the false feedback in the first round. Dyads were most likely to change their preference for non-detected manipulations, where their preference changed in 34.7% of trials. Response times and choice confidence during inconsistent non-detected manipulated trials matched those found in consistent non-detected trials, a pattern which was not found for detected manipulated trials or non-manipulated trials.

Paper III – How information availability interacts with visual attention during judgment and decision tasks.

Research question

When making a decision in front of a supermarket shelf, consumers are likely faced with a mix of visually available information and associated memories – as well as interactions between those two. Cognitive processes, such as decision-making, search, and various judgments are therefore likely to co-occur, and each process will have its impact on visual attention. To better understand how visual availability of information shapes the deployment of visual attention, we contrasted the use of gaze during decision and judgments when facing task environments with and without task-relevant information.

Procedure

In the first experiment, we recruited sixty-three participants and used a remote eye-tracker to study their visual attention. Participants completed forty trials which each consisted of three novel product options presented during a fifteen second long encoding phase. Each option consisted of three attributes and their values, for example price, sugar content, fruit content. Following encoding, participants were asked to either choose which option they preferred or make a judgment about one of the options. Half the time, participants viewed the same environment as during encoding and half the time all the attribute information was removed.

The second experiment replicated the first with thirty-eight participants, with the only difference that when attribute information was removed it was replaced by symbols ('###') instead of being completely blanked out.

Results and Conclusion

In both experiments we found that participants' visual attention during decisions were sensitive to evaluations made already during the encoding phase. When no task-relevant information was present participants appeared to search for, and quickly orient to, their preferred option. When the task environment allowed, participants engaged with it using slow integrative processes to improve the quality of both their decisions and judgments. The findings have implications for how to best study visual attention and decision-making in ecologically valid ways.

Paper IV – I see your dilemma: Visual attention and moral choice.

Research question

Trolley problems have long been used to ground theorising about human moral psychology. In this paper, we used novel graphic representations of Trolley-type dilemmas to investigate participants' decision processes. Of particular interest was to compare participants' gaze patterns when responding with the deontological option or the utilitarian option, as these types of moral choices have been hypothesised to arise from different underlying psychological processes.

Procedure

We studied 55 participants responding to fourteen trolley-type moral dilemmas. Each dilemma was represented as two rows of two-panel cartoon strips. Each row corresponded to either a utilitarian or a deontological response to the dilemma. In the first panel of each row, an action was presented, and in the second panel the outcome of that action was shown. Participants were asked to select the row which they found morally right. Participants' eye-movements were recorded using a tower-mounted eye-tracker.

Results and Conclusion

Participants displayed general gaze biases consistently found for other decision-making tasks. When choosing the utilitarian option, participants showed higher transition frequencies, less deterministic gaze patterns and spent more time looking towards the outcome panel of the chosen option. The findings give support to accounts emphasising that utilitarian choices are made by more rational, cognitive processes compared to deontological ones.

Paper V – Biasing moral decisions by exploiting the dynamics of eye gaze.

Research question

Eye gaze and other sensorimotor processes have been hypothesised to track decision trajectories as they unfold. This paper investigated if it is possible to change the course of an individual's moral decision using information derived from gaze alone.

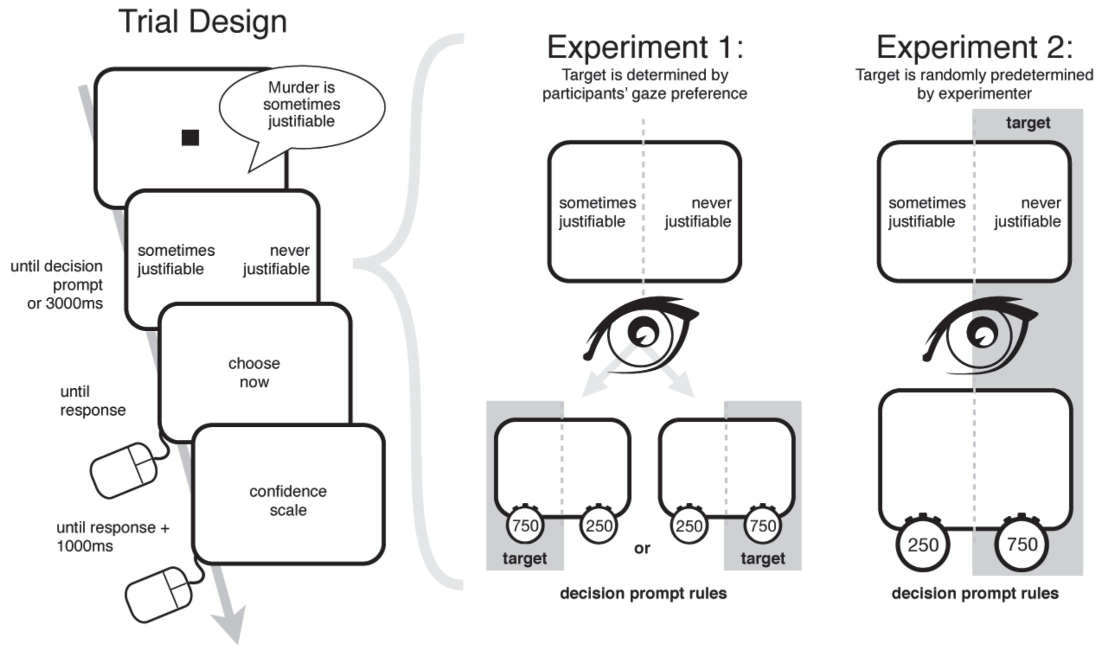

Figure 1. Illustration of the gaze-contingent interruption paradigm introduced in Paper V. Participants view alternatives and their decisions are prompted when they have distributed their gaze according to predefined rules. In Experiment 1, participants had to have viewed any one alternative at least 750ms and the other alternative at least 250ms for their decision to be prompted. In Experiment 2, one alternative was randomly predetermined as target, and participants' decisions were prompted once they had viewed the target alternative for at least 750ms and the non-target for at least 250ms.

Procedure

In three experiments, we introduced a novel gaze-contingent interruption paradigm. Participants freely viewed response alternatives to abstract moral statements such as "Is murder justifiable?" Concurrently, their eye-movements were monitored using a remote eye-tracker and participants' total dwell time to each alternative aggregated. Once a pre-defined amount of dwell time to each alternative had been reached, participants' deliberation was interrupted and their decision prompted. In the first experiment participants' decision was prompted as soon as one alternative reached at least 750ms of dwell time and the other at least 250ms. The alternative viewed the longest was designated as 'target'. In the second and third experiments the target was *randomly assigned prior to trial onset* and participants decisions were prompted once they had viewed the target for at least 750ms and the other alternative for at least 250ms (see Fig. 1).

Results and Conclusion

We found that participants would choose the target alternative in 59.6% of trials in Experiment 1, indicating that eye gaze tracked the developing moral decision process. In Experiments 2 & 3, participants chose the randomly assigned target alternative in 58.2% and 55.4% of trials, demonstrating that participants' moral choice could be biased by manipulating the timing of their decision based on monitoring eye gaze. Further analysis showed that the current direction of participants' eye gaze was more important for determining their choice compared to relative exposure, which might indicate the presence of a leaky integrator underlying evidence accumulation in this task.

Paper VI – Modelling moral choice as a diffusion process dependent on visual fixations.

Research question

In this paper we investigated if an influential computational decision model utilising visual attention, the attentional drift-diffusion model (aDDM), could be used to fit data on moral choices between abstract alternatives.

Procedure

Eighteen participants responded to a two-alternative forced choice task (2-AFC) consisting of sixty-three moral decisions. In each trial, participants first heard a moral statement, e.g. "Hurting a defenceless animal is one of the worst things one can do", and then two alternatives were presented, e.g. "It's always bad", "It's sometimes bad". While participants decided their eye-movements were recorded using a tower-mounted eye-tracker. After each trial, participants also rated the relative goodness of the chosen option compared to the non-chosen option. Three models were tested; two models took fixations into account but differed in how they assigned the relative goodness value to the two options. The third control model was a standard diffusion model which did not take fixations into account.

Results and Conclusion

The results show that a model taking visual attention into account outperforms a model that does not. However, none of the models could fully account for fixation properties of the decision process observed. Still, the overall results indicate that moral choices can fruitfully be modelled as diffusion processes, highlighting the possibility for developing domain general computation decision models.

Paper VII – A fixation dependent decision model of charitable choice.

Research question

The aim of this paper was to investigate the computational properties of decision between charitable organisations in the context of an aDDM approach.

Procedure

Twenty six participants first rated 31 charities on how praiseworthy they considered their work to be. Participants were then asked to choose between random pairing of these organisations in a 2-AFC. Participants made a total of one hundred choices. While deciding participants eye-movements were monitored in a tower-mounted eye-tracker. Fixation dependent and fixation independent versions of the aDDM were then fit to the data.

Results and Conclusion

A fixation dependent model was found to provide the best fit to the response time data. Additionally it could account for response times, choice distributions, and many other aspects of the participants' fixation behaviour. The results indicate that evidence accumulation driven by visual attention underlies choices in the moral domain.

Dynamics, body and attention

> It's like in those old maps of the world, where mapmakers wrote, "Here be dragons" on the unexplored parts of the globe. These topics are not completely unexplored, of course, but it is fair to say that they lie at the limits of current understanding. The problems are very hard, because they are both large and nonlinear. The resulting behaviour is typically complicated in *both space and time*, as in the motion of a turbulent fluid or the patterns of electrical activity in a fibrillating heart. – S. Strogatz (1994, p. 11)

This thesis is grounded in a dynamic perspective, whereby cognition is to be understood as fundamentally continuous and embodied. The following chapter outlines this view and discusses experimental evidence related to using eye gaze and mouse arm movements to understanding cognition.

From symbols to dynamics

> First, I maintain that *the mind*, which we often call the intellect, the seat of the guidance and control of life, *is part of man*, no less than hand or foot or eyes are parts of a whole living creature. – Lucretius (2005, p. 12)

As a reaction to the dogmatic behaviourism of post-war America, cognitive science emerged in the 1950s and 60s as a result of conceptual breakthroughs, technological developments, and bright people finding themselves in the same room (Miller, 2003; Gärdenfors, 2008). The new science focused on information and information processing, and since computers seemed to do much the same thing, the language of computers and computing provided the new metaphors of the mind (e.g. Atkinson & Shiffrin, 1968). Cognition, and especially higher-order cognition – i.e. language, categorisation, problem-solving and memory – needed representations to operate on, and these were thought to be symbolically and sequentially manipulated (Newell & Simon, 1976; Marr, 1982; Pylyshyn, 1984). In the sections that follow, I instead focus on an alternative way of understanding the mind beyond the symbolic-computational approach, grown out of both its shortcomings as well as the development of new analytical and metaphorical tools.

Despite the symbolic approach's close affinity with computer science and conventional AI, it turned out that creating intelligent, flexible systems mimicking human abilities was slightly more difficult than had been originally supposed (Minsky,1974). Instead, it was suggested that as an alternative to building systems based on symbolic manipulations of abstract concepts, with intelligence 'built-in', a better method would be to allow intelligence to be discovered through piecemeal interactions of the systems with its surroundings (Harnad, 1990; Brooks, 1991; Ballard, 1991). Doing so does away with symbolic representation as a precondition for cognition, but still allows for goal-like and planned – *intelligent* – behaviour as an emergent property of the system (e.g. Balkenius, 1995). Another important development was the connectionist approach to computation, spearheaded by the Parallel Distributed Processing Research Group (PDP; see Rogers & McClelland, 2014 for an overview). This framework emphasises how interactions between simple, non-symbolic parts can lead to emergent, complex behaviours and cognition. The modelling tool of connectionism was the neural network in both feed forward and recurrent configurations (e.g. Hopfield, 1982). What connectionism entailed was a view of cognition as a process unfolding in real-time, represented by activations among weighted connections between neuron-like units. Studying how activations or connection weights changed in a neural network would also allow

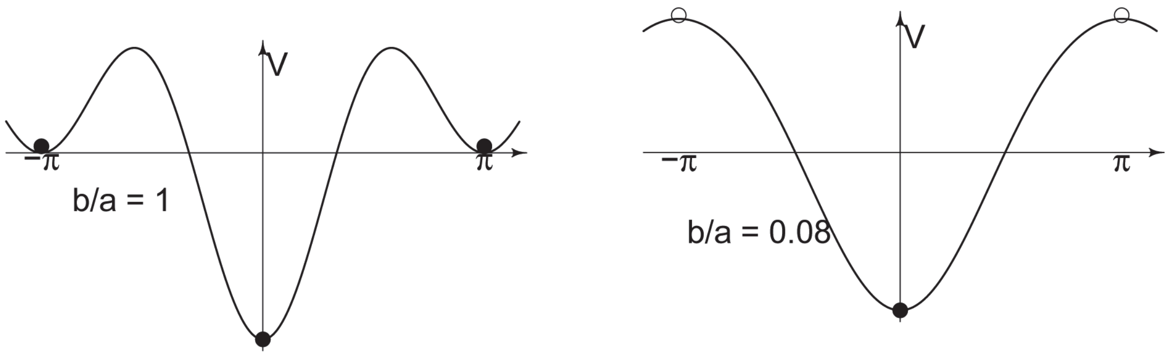

Figure 2. Illustration of the HKB model for two different ratios of *b/a*. Black circles indicated stable states. White circles indicate unstable states, from which the system will move.

the researchers to study the underlying, modelled, cognitive process. Many of the early neural networks and their learning methods were far from accurate as biological models (Rumelhart, Hinton & Williams, 1986), nevertheless the connectionist approach probably came much closer to proposing a plausible picture of cognition compared to the symbolic approach.

While the information processing perspective borrowed from engineering and the connectionists from neurobiology, the third major view on cognition, the dynamical systems approach, took its' cue from physics and the mathematical tools developed to study complex systems (Van Gelder & Port, 1995; Kelso, 1995; Smith, 2005; Schöner, 2008). Many of the insights of the connectionist framework can be recognised, for example in the idea that competition among dynamical systems (neurons, bodies, individuals) leads to emergent behaviours and outcomes (Grossberg, 1980b; Cisek & Kalaska, 2005; Wong & Wang, 2006; Elman, 1991). Dynamical systems models typically approach cognition from a more general standpoint, emphasising the similarity in models across fields. From this perspective, the same concepts of competition and emergence that are found in neural network modelling can also be applied in the study of natural phenomena from piles of sand to evolutionary population dynamics (Grossberg, 1980a, 1980b; Bak, 1996). An example is the Haken-Kelso-Bunz (HKB) model of coordinated, rhythmic finger movement (Haken, Kelso & Bunz, 1985; Kelso, 1995). The experimental situation that is being modelled is the case of moving one's index fingers back and forth to the tick of a metronome. The model, at its simplest, is defined by the potential function V:

$$V = -a\cos(\varphi) - b\cos(2\varphi)$$

Where φ captures the relative phase of the system, in this case the relative position of the fingers, which can either be in phase (φ = 0) or antiphase (φ = $\pm\pi$) to each other. The variables *a* and *b*, or rather their ratio *b*/*a*, capture period of rhythmical coordination – in experiments the beat of the metronome – with decreasing ratios indicating shorter periods (faster movements). Figure 2 illustrates the landscape generated by *V* for two ratios of *b*/*a*, showing the transition from bistable states, i.e. allowing both in-phase and antiphase movements, to only having one stable state, the in-phase movement. By analysing this system, further empirical predictions can be tested. The basic model, due to its tractable dynamics, has also been applied well beyond simple motor movement, e.g. to speech categorisation (Tuller, Case, Ding & Kelso, 1994) and perspective taking (Duran & Dale, 2014). What this excursion illustrates is the seemingly radically different way in which the dynamical systems approach attempts to conceptualise and model cognition.

Tracking continuous cognition

> Dynamical systems accounts of mind are often perceived as foreign objects in the body of psychology. They are poorly understood, and if their descriptions aren't annoyingly vague, then their math is daunting. Perhaps most problematic is the simple fact that for some scientists, it simply conflicts with introspection to claim that the mind does not think one discrete "thing" and then think another discrete "thing." – M. Spivey (2007, p. 80)

The papers in this thesis do not take a strict dynamical systems approach, helping themselves to the full smorgasbord of nonlinear modelling available. Instead, they make use of two important insights about cognition that underpin the dynamical systems view; that cognition is continuous and embodied (Spivey, 2007; Kelso, 1995; Gibson, 1986).

The term "embodied cognition" has many uses, so a clarification is in order to avoid misunderstanding. One common view is that sensorimotor representations are partly constitutive of conceptual representations, a view we can call metaphorical or analogical embodiment (Lakoff & Johnson, 1980; Wilson, 2002). This is an idea underpinning a lot of recent research, for example, in moral cognition (e.g. Schnall, Benton & Harvey, 2008; Zhong & Liljenquist, 2006). On this view, some aspect of experience, i.e. an object being hot vs. cold, clean vs. dirty, etc., is hypothesised to influence and give evaluative meaning to our higher order conceptual representations by grounding them in analogical sensorimotor experiences. For example, our knowledge of hotness is constitutive of our knowledge of goodness, therefore holding a hot cup of coffee will make us evaluate something in a more positive manner.

In contrast, the embodied view used in this thesis takes as its starting point the observation that cognition is fundamentally for action in an environment (Wilson & Golonka, 2013; Spivey, 2007; Gibson, 1986). Sensorimotor processes reflect ongoing cognitive activity due to the continuous and distributed nature of cognitive processing.

To see what this means, consider an example. Imagine the act of picking up a glass of Kool-Aid and drinking it. On a symbolic computing account, the executive part of the brain would decide to pick up the glass and, consequently, send a command to the motor part of the brain. While the actual arm movements might have their inherent dynamics, this would be considered to be separate from the actual execution of the command. Instead, now imagine closely observing someone while they are about to begin reaching for the Kool-Aid. There are several things that you would possibly observe. For example, their fingers might twitch a bit or maybe their arm might startle back and forth before clearly moving towards the glass. Once the arm is 'in flight' maybe sometimes it goes straight for the Kool-Aid and maybe sometimes it instead seems to veer towards a nearby Piña Colada before reaching its intended destination.

The continuity claim is that all these movement differences matter for understanding cognition. That is, they reveal information not only about the motor aspect of reaching for the Kool-Aid, but also about the underlying decision to initiate the motor command in the first place. They do so because cognitive states are continuous between themselves; i.e. that neural computations do not operate serially. Instead the areas in the brain responsible for executing an arm movement are continuously fed information from areas responsible for calculating that reaching for the drink is appropriate, where the location of the target is, and so forth. Hence, the sensorimotor system will continuously reflect the 'outputs' of these classically cognitive computations as they occur in real time (Bastian, Schöner & Riehle, 2003; Cisek & Kalaska, 2005; Gold & Shadlen, 2000; Zelinsky, Rao, Hayhoe & Ballard, 1997; Spivey, 2007).

To take another example, consider categorising the object depicted in Figure 3. Before, for instance, judging it to be a duck and not a rabbit, one will literally be in between both judgments while the cognitive system evolves as a function of the probabilistic nature of neuronal computation (Grossberg, 1980b; Pouget, Dayan & Zemel, 2003). Moreover, the judgment one ends up having is likely not a 'pure' state either; rather it is also a probabilistic state reflecting whatever activation levels the respective judgments of 'duck' and 'rabbit' had at the time the judgment was pronounced. To see why, consider the following: neuronal computation works so as to achieve activation in relation to some threshold (Hanes & Schall, 1996; Gold & Shadlen, 2001), but the threshold can be variable (Kiani, Hanks & Shadlen, 2008). This means that even if a one-to-one mapping of mental states to the activities of neuronal assemblies could be

Figure 3. Ambiguous figure of rabbit and duck.

achieved, the states could be constituted by varying levels of activity due to varying thresholds. Furthermore, the mind is always literally in motion; once one has attended Figure 3 long enough to make up one's mind, attention will be directed elsewhere and the state of one's duck-judgment will change. This is sometimes also phrased to say that cognition is fundamentally probabilistic and graded (Spivey, 2007).

Early evidence for continuous and embodied cognition in tasks beyond motor control was found in linguistics with the development of what is now known as the Visual World Paradigm (VWP; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Kamide, Altman & Haywood, 2003; Andersson, Ferreira & Henderson, 2011). In the VWP, participants are asked to view a scene while listening to a sentence, which usually describes some or all of the objects present on the screen. Participants' eye-movements are measured while they listen to the sentence. In the defining study of the VWP, participants were asked to watch a display containing two towels, an open box and an apple placed on one of the towels. Comparing participants' gaze trajectories when hearing target sentences like "put the apple on the towel in the box" compared to "put the apple that's on the towel in the box", showed radical differences in both when, and how long, participants looked at the depicted objects. This indicated that eye-movements were sensitive to immediate processing differences of subtle syntactic features of the target sentences (Tanenhaus et al, 1995).

Similarly, investigating the time course of participants' gaze using the VWP has been shown to reveal how interpretations of others' mental states can arise early during understanding, indicating that theory of mind inferences can be performed remarkably fast (Ferguson & Breheny, 2011). Eye-movements have also been used to study recall in memory tasks. One prominent finding there is that people are more likely to look towards areas where information they are trying to recall previously was presented, even if that area now contains no information at all (Richardson & Spivey, 2000).

Recent work has demonstrated that eye-movements might not only reflect ongoing cognition processes but also facilitate them. Participants' recall was compared when directed to look towards a spatially congruent portion of the screen or when directed to look towards an incongruent portion in relation to the memory probe. Using this manipulation, Johansson and Johansson (2014) showed that recall and response times were facilitated by congruent gaze manipulations. Similarly, there is evidence of a facilitating role of eye gaze during insight problem learning tasks. By directing participants' gaze to portions of a visually presented problem where people solving the task tend to look, the success rate of the problem was doubled (Grant & Spivey, 2003).

By measuring arm movements, typically by tracking computer mouse trajectories, similar dynamics have been uncovered as when tracking gaze. For example, participants were instructed to respond to the utterance "candle" by clicking the object the word refers to out of two objects, one target and one distractor, shown onscreen. When the distractor object was phonologically similar to the target, such as "candy", there was more curvature in the mouse movement compared to when the distractor object was phonologically dissimilar, such as "jacket". The curvature in the resulting mouse movement can be understood as arising from competition between the alternatives from the linguistic processing of the utterance. This influences the programming of the arm movement concurrent with participants responding to the task (Spivey, Grosjean & Knoblich, 2005)[2]. Using mouse movements to measure how continuous processing evolve between competing alternatives has also been implicated in sex-based facial categorising (Freeman, Ambady, Rule & Johnson, 2008), as well as, for perspective taking in social judgments (Duran & Dale, 2014) and for simple truth decisions (McKinstry, Dale & Spivey, 2008).

---

[2] This is the experimental analogue of the arm veering close to the Piña Colada while reaching for the Kool-Aid earlier.

There is thus considerable evidence for the usefulness and empirical adequacy of the overarching theoretical view outlined above. In addition to these empirical claims, one important reason for taking this embodied and continuous view is that it provides a powerful framework for hypothesis generation and for organising results and data. When taking embodied dynamics into account, experiments give more data: data that arguably contain a higher resolution snapshot of the cognitive process of interest. This is the difference between just analysing response times to a task (*pace* Sternberg, 1969), and also analysing what was happening during the response. Having said this, it is important to make clear what I am not claiming – that this particular theory is the final view on human cognition. The project of understanding and explaining cognition can be formulated at so many levels that an eclectic and pluralistic approach is necessary (cf. Dale, 2008).

The papers in this thesis reflect the embodied and continuous perspective discussed in different ways. In Papers I, III and IV the time course eye gaze is used to investigate aspects of preference formation and decision-making, using task relevant displays in a manner building on the spirit of the VWP. In Paper V the idea that cognitive states are probabilistically reflected in sensorimotor activations used to bias moral choices. Papers VI and VII develop computational models of the time course of eye-movements during choice tasks.

The next chapter develops a view on preferences and discusses eye movements in relation to choice and preference formation, while the next one again relates eye movements to moral cognition.

Interlude: On gaze and attention

> … natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system. – J.J. Gibson (1986, p. 1)

It is common place to simply assume that eye gaze is to be taken as equivalent to visual attention. However, it is relatively easy to empirically show that they are not the same thing; for example, by asking participants to maintain fixed gaze while performing a peripheral discrimination task, attention is readily dissociated from gaze direction (Posner, Snyder & Davidson, 1980). Despite this, much of the work studying eye gaze to understand underlying continuous processes takes this assumption for granted. To see why, first consider the extreme alternative case, that eye gaze is unrelated to visual attention; that is, it is completely random with respect to underlying cognitive processes. Eye fixation patterns during a visual search task exhibit $1/f$ noise; this indicates long correlations in the temporal structure of fixations that are signatures of non-random dynamic processes (Aks, Zelinsky & Sprott, 2002). Furthermore, when comparing human visual search patterns to those predicted from an ideal Bayesian observer, human observers perform nearly optimally with regards to saccade selection (Najemnik & Geisler, 2005). This indicates that humans maximise the information gain from the environment through the dispersion of fixations, which is the most efficient strategy given that the human visual system exhibits imperfect information integration between fixations - as arguably change blindness shows (Rensink, O'Regan & Clark, 1997; Najemnik & Geisler, 2005). The relation between gaze and attention is, likely, highly task dependent. For example, gaze direction does not predict detection in a change blindness task unless the current task demands require processing of precisely those which are changed (Triesch, Ballard, Hayhoe & Sullivan, 2003).

Additionally, there is considerable evidence that while it might be possible to dissociate attention by shifting it without a corresponding eye-movement, the opposite, i.e. shifting gaze without also switching attention, is not possible. Studies on macaque monkeys have shown that attentional systems share their neural substrate with eye-movement planning in the superior colliculus (Kustov & Robinson, 1996). Data on humans performing visual search tasks also suggests that attention and saccade targets are closely coupled (Deubel & Schneider, 1996). This might result from competition amongst objects in the visual field for the limited and privileged resources associated with full foveation (Desimone & Duncan, 1995; Kukona & Tabor, 2011).

The strong eye-mind hypothesis formulated by Just and Carpenter (1976) suggests that the direction of gaze reflects what the cognitive system is currently processing. By measuring fixation direction and fixation duration, they argued that it is possible to investigate both what and how consuming current mental operations are. While this is likely too strong an assumption to hold, a qualified version of it seems plausible in light of the evidence reviewed above.

In sum, given constrained experimental designs, attention and gaze can be treated as co-occurring, and thus, informative about cognitive processes in general.

On preference and choice

> ET's spaceship was waiting for him. As he boarded the craft, he mused that the report to home base would be easy. 'The earthlings are stuck in a quagmire. They don't see that brains are decision-making devices and should be understood in those terms – that level of description, not lower. They are only partially evolved. It will be eons before they ever find us. It might also be eons before they ever understand themselves.' – M.S. Gazzaniga (2010, p. 292)

In this chapter I begin by considering the questions of preference as a measurement problem, starting with a classical view from economics, and reviewing challenges from psychology from the 1950s to the present day. The aim will be to ground a view of preferences as being dynamically constructed, and to relate this view to current work on visual attention and choice.

Revealing preference

> Woe to any who deny any one of the three postulates here! – P.A. Samuelson (1938, p. 70)

Choices are thought to be explained by agents' underlying preferences; preferences are theorised mental states, and as such not directly amenable to observation. To solve this, economists have suggested that observable choices can be used to infer the preferences of the agent – this is the idea of revealed preferences (Samuelson, 1938; 1948; Edwards, 1954, Grüne, 2004).

To simplify, revealed preference theory states that, given some choices and a budgetary constraint, if an agent chooses some bundle of goods x_i while she could have afforded some other bundle x_j, then x_i is revealed to be preferred to x_j[3]. In other words, preferences are given by agents' actual choices. However, ideally, what behavioural scientists[4] want to understand is how an agent faced with two prospects goes about choosing between them, based on her preferences. Given this, it might seem a bit backward to go from choices to preferences and then back again. However, it is not quite so awry if one considers two things: (a) revealed preference theory is embedded in an axiomatic approach to decision-making, allowing the theorists to derive a number of interesting things *if* something like the presuppositions behind revealed preferences hold, and, (b) if one assumes that agents quite straightforwardly can maximise among their preferences, then knowing the preference ordering is all that is required to predict their future choices.

The important lesson from this is that through the framework of revealed preferences, as well as the related and more general expected utility account of choice, we find assumptions of complete and ordered (transitive) preferences readily available to the agent (Edwards, 1954; Glimcher, 2010). However, preferences have been found to be less well-behaved than classic economic and decision theory would expect. For instance, when giving participants a choice between a gamble involving a high probability of winning a small amount of money (P bet) versus a gamble involving a low probability of winning a large amount of money ($ bet), participants will often select the P bet. However, when asked to instead bid for the two gambles, participants will often assign a higher monetary value to the $ bet. This robust phenomenon is known as preference reversal (Lichtenstein & Slovic, 1971; Slovic & Lichtenstein, 1983). A second demonstration of the misbehaviour of preferences is that of framing (Tversky & Kahneman, 1981; Petrinovich & O'Neill, 1996). Decision frames refer broadly to the agent's understanding of the decision situation deriving from the wording used to pose the decision problem. In the most famous example, the Asian Disease Problem, participants are asked to choose between policies for dealing with an expected outbreak of a particularly virulent strand of

---

[3] This is the weakest form of revealed preferences; preferences can also be strictly revealed and indirectly revealed.

[4] I use the term inclusively to include both psychologists, economists, sociologists, cognitive and neuroscientists.

an Asian disease [i.e. the flu] which is expected to kill 600 people. In one version of the problem, participants can choose between program A which will save 200 people for sure or program B which will save 600 people with 2/3 probability. Most participants choose A. In a second version, participants choose between program C under which 400 people will die for sure and program D under which 600 people will die with 1/3 probability. Most participants choose D. However, the problems are mathematically identical, as the expected outcome of options A/C and B/D are the same. It appears that decision makers respond differently to loss versus gain framings; i.e. the framing shifts the perspective of the agent and affects what values she assigns to the decision outcomes.

More generally, these findings of procedural and descriptive variance, meaning that the method and formulation when preferences are elicited affects what preferences are revealed, have been taken to imply that general information processing limitations of human cognition also apply to decision-making (Tversky & Kahneman, 1981; Slovic & Lichtenstein, 1983, March, 1978). More generally, this is the bounded rational approach to human cognition, which implies that agents instead of maximising engage in 'satisficing' (Simon, 1956); i.e. that agents aim to do "well enough" relative to their environment and their cognitive abilities. This suggests a situational approach to preference and choice. The agent's cognitive abilities determine, in tandem with the constraints of the decision situation, what decision strategy is used and which the agent's preferences are. Both prospect theory (Kahneman & Tversky, 1979) and the adaptive decision maker approach (Payne, Bettman & Johnson, 1993; Gigerenzer & Goldstein, 1996) are examples of traditions in this vein.

Another striking demonstration of the situational aspects of choice and preference is *choice blindness*, introduced by a group of Lund University Cognitive Science (LUCS) researchers (Johansson, Hall, Sikström & Olsson, 2005). As this has interesting and important implications for my understanding of preferences and Papers I and II are written within the choice blindness paradigm, I will discuss choice blindness at some length.

The challenge of choice blindness

> If you reach for your car keys, you don't end up with an armadillo in your lap. – L. Hall

Choice blindness is the finding that participants, following a choice, are willing to accept their non-chosen alternative as the outcome of their choice when given false feedback about the choice outcome. In the original study (Johansson et al., 2005), participants made choices between pairs of female faces. The faces were presented on cards, which the experimenter held in his hands. Following each choice the experimenter would lower the cards to the table, and then slide the chosen card towards the participant in one continuous movement. By using a technique adapted from stage magic participants would, on manipulated trials, be presented with the card opposite of what they chose. Not only did participants fail to detect the manipulation in a majority of trials, but they would also, when prompted to state reasons for their choice, proceed to confabulate such reasons. These confabulations have in several studies been found to be remarkably similar to reasons given to actual, non-manipulated choices (Johansson et al., 2005; Johansson, Hall, Sikström, Tärning & Lind, 2006; Hall, Johansson, & Strandberg, 2012). These findings highlight and underscore previous seminal work by Nisbett and Wilson (1977) showing that being able to generate a verbal report about a mental process does not imply veridical introspective access to that process. In one oft-cited demonstration, Nisbett and Wilsson asked participants to choose between four identical pairs of nylon stockings presented on a rack. There was a strong right-most bias in the choice data, but none of the verbal reports by the participants factored in position as a reason for their choice. While some of these original findings have been criticised on methodological grounds (see White, 1980), the discovery of choice blindness clearly vindicates the general lesson about introspective unreliability for process reports.

Choice blindness has been demonstrated for a host of different choices and domains. Agents can be found to be choice blind for choices between abstract artistic patterns (Johansson, Hall & Sikström, 2008), financial decisions (McLaughlin & Somerville, 2013), eye-witness identifications (Sagana, Sauerland & Merckelbach, 2013), as well as for difficult moral (Hall, Johansson & Strandberg, 2012) and political judgments (Hall et al., 2013). The effect has been shown for modalities as different as taste and smell (Hall, Johansson, Tärning, Sikström & Deutgen, 2010), touch (Steenfeldt-Kristensen, & Thornton, 2013) and auditory presentation (Sauerland, Sagana & Otgaar, 2012). Paper II demonstrates choice blindness for two collaborating participants for the first time. Even after the process of explicitly negotiating and agreeing on a mutual choice, was the false feedback about the outcome of the decision accepted by the dyads.

It could have been the case that in the original experiment, and in the many replications and extensions, participants refrain from voicing concerns about the manipulation despite having been aware of it. They could be doing so perhaps due to embarrassment towards the experimenter or to minimize experienced dissonance towards the presented choice (Festinger, 1957). In Paper I, participants' responses to the false feedback during a choice blindness task was investigated using their eye-movements and by measuring their pupil dilation. The results indicated that detected trials were associated with both processing differences and increased pupil signals compared to non-detected trials. Furthermore, the results indicated that, qualitatively, non-detected trials shared many similarities with non-manipulated trials (see Paper I for details). The results, together with the fact that participants in the version of the choice blindness task used in Paper I sat alone with a computer with no direct experimenter interaction, help rule out much of the aforementioned misgivings.

Importantly, investigating the downstream effects of choice blindness manipulations have shown that accepting the manipulations has pronounced effects on later choices and attitudes of participants. In a study conducted during the run-up to the 2010 Swedish general election (Hall et al., 2013), participants were asked to rate their voting intentions prior to indicating their agreement with a number of statements concerning political questions dividing the right- and left-wing coalitions. Participants were subsequently given false feedback about these ratings and their resulting coalitional alignment (which was calculated from their responses to the specific political questions). Finally, participants were asked to again give their voting intention. Participants significantly shifted their voting intentions, with almost half the participants being willing to consider a coalitional switch. In fact, a full 10% of the participants completely shifted coalitional allegiance in their stated voting intentions. Similarly, in another experiment (Johansson, Hall, Tärning, Sikström & Chater, 2014; see also Taya, Gupta, Farber, & Mullette-Gilman, 2014), participants were asked to make ratings of faces prior to a choice blindness phase identical to the one in the original 2005 study. Following the manipulations, they were asked to rate the faces a second time as well as to perform a second round of choices. Participants rated the believed-to-be chosen faces much higher and preferred the believed-to-be chosen face in over half the trials. People appear thus not only to be blind to the outcomes of their choices, but also very sensitive to false feedback, to the extent that this will affect their future preferences.

What then do findings from the choice blindness literature tell us about preference formation? First of all, it is important to note what the findings do not imply. Tempting as it may seem, choice blindness does not show that people do not have stable preferences at all or are never able to act on them. Instead, findings of choice blindness highlight the importance of the immediate environment for the ongoing cognitive processes of the agent. We continuously respond to our environment and what is in it, having evolved, as a species, and grown up, as individuals, in a relatively stable world where things do not suddenly change places (O'Regan & Noë, 2001, Hall & Johansson, 2008). The implication for preferences is clear: like the rest of the mind, they too are constituted and constructed over time. There is no need to posit a complete set of preferences for the sake of the agent's decision-making capability – after all she has the world to rely on and "the world is its own best model" (Brooks, 1991). In other words, choice blindness points out a route for us to circumvent some of the motivations for the assumption of complete and stable preferences, by way of the ordinary stability of the external world.

Choice blindness also indicates that any proposed links between intentions and choices should be treated with utmost caution. In the field of motor control, numerous models have been suggested whereby action outcomes are continuously monitored and compared with a prior intention underlying the action specification (Wolpert, 2007). The success of these models in low-level motor control has led to suggestions that similar architecture might be a pervasive feature of human cognition. For example, comparison with prior intentions has been proposed to underlie our sense of agency (Haggard, Clark & Kalogeras, 2002). If such a mechanism is posited for decisions, then choice blindness indicates a disquieting failure of such monitoring. Given recent work using a real-time speech exchange method to give false-feedback of speech production (Lind, Hall, Breidegaard, Balkenius & Johansson, 2014), another area where strong intention monitoring has been posited, any strong comparator approach to choices is likely off the table on the current state of evidence (cf. Paper I).

Processing constructed preferences

> We construct our preferences. We choose preferences and actions jointly, in part, to discover – or construct – new preferences that are currently unknown. – J.G March (1978, p. 596)

Given a bounded rational approach to cognition and the findings of preference anomalies in relation to standard decision theory, an alternative view of preferences as being constructed suggests itself (March, 1978; Slovic, 1995; Payne, Bettman & Schkade, 1999, Ariely & Norton, 2008). On this view preference elicitation triggers a strategic reaction in the agent, leading her to construct her preferences in relation to the information that is presented and the demands of the situation. This entails that the preference does not exist independently of the choice situation. Strongly put, there was nothing there to be measured until the agent was queried (Slovic, 1995)! Of course not everyone has accepted this particular conclusion. According to the discovered preference hypothesis (Plott, 1996, Braga & Starmer, 2005), agents do have stable and context free preferences; however, these are not always immediately accessible to the agent. Instead she *discovers* her preferences over the course of familiarizing herself with the decision environment and, in essence, with *herself*. The main differences between these two views are ultimately ontological – it is a question of what kind of entities we want to assume to exist in our minds (Fischhoff, 1991). On a discovered preference view, convergence towards optimal decisions over time (optimal according to expected utility theory) in experimental situations is evidence that the agent is discovering her true, stable preferences. That is, violations of normative theory occur only *before* preference discovery.

It is instructive to compare these views to another research area in the psychology of choice and preference, one that is often omitted from the more economically oriented literature. I am referring to the effect that is known as preference change through choice and the associated Free Choice Paradigm (FCP). Preference change through choice was first reported by Brehm (1956) in a study utilising the FCP. In that study, participants were asked to rate a number of household items on how desirable they were. Following the rating, participants were asked to choose between two of the items and were told they would receive the one they chose. Finally they were asked to again rate the items. Chosen items were rated higher and non-chosen items were rated lower, compared to original ratings, in the second rating phase. The differences were larger for closely rated items. Thus simply making a choice can affect one's future preferences in relation to the options chosen between (cf. Morwitz, Johnson & Schmittlein, 1993 for evidence that the mere *measurement* of preferences also affects future choices).

Preference change through choice was originally interpreted using the framework of cognitive dissonance theory (Festinger, 1957; Harmon-Jones & Harmon-Jones, 2002). According to this framework, agents seek consistency between outcomes in the world and internal cognition. When inconsistencies are detected, agents experience dissonance, a not necessarily conscious, mental state which, on some level, can be considered "unpleasant" for the cognitive system. As a result, they seek to reduce it by changing the underlying attitudes or cognitions creating the inconsistency with the outer world[5]. For example, in a study by Festinger and Carlsmith (1959) participants performed a long and boring experimental task and were then paid either $1 or $20 to (untruthfully) tell a fellow student waiting to participate that the experimental task was, in fact, enjoyable. Participants were then asked how enjoyable they themselves had found the experimental task. The main finding was that participants in the $1 condition rated the task as significantly more enjoyable. This was interpreted as resulting from participants in the $1 condition experiencing much more dissonance between their original cognition (the task is boring) and deceiving the fellow student in relation to the compensation.

An alternative account of this type of preference change is that of self-perception theory (Bem, 1965; 1972). On this account, grounded in a functional, behaviourist perspective (e.g. Skinner, 1948), agents infer their own preferences by observing their own behaviour (choices) much in the same way they would infer others' preferences by observing their behaviour. This account does away with the needing to assume a specific cognitive state (i.e. dissonance) to explain the preference change through choice effect. On those grounds, self-perception theory might be considered to be the more parsimonious account.

Recent experiments investigating preference change through choice have, in response to severe methodological criticisms of the FCP (Chen & Risen, 2010),

[5] Curiously, agents do not seem motivated to enact further changes (for example, by making choices) to the world to alleviate cognitive dissonance.

provided new evidence consonant with a self-perception account. In one experiment, participants were asked to make choices between purportedly subliminally presented options (vacation locations). In fact, only gibberish was presented resulting in that participants made choices independent of any preferences they might have had. Participants rated these blindly chosen options higher than the non-chosen options post-choice compared to pre-choice (Sharot, Velasquez & Dolan, 2010; cf. Egan, Santos & Bloom, 2010). The choice blindness studies previously reviewed also operate on the level of participants' beliefs by directly manipulating these through false feedback (Johansson et al., 2014; Taya et al., 2014).

The findings in Paper II are particularly instructive concerning the effects of beliefs on future choices. In that study pairs of participants formed dyads and were instructed to make mutual choices between faces presented in pairs. On some trials, dyads were given false feedback regarding their choice. In a later stage of the experiment, dyads were asked to make a second round of choices between some of the same face pairs as previously. Dyads were as consistent as individuals for non-manipulated trials, but often changed their preferences for manipulated trials, with the effect being largest during non-detected trials. Importantly, in the case of dyads there is little reason to believe that there was something like a mutual preference waiting to be discovered or revealed. Instead, the crucial mediating factor appears to be participants' beliefs about their previous choices and actions. This suggests that a self-perception account might be able to better countenance the evidence from the preference change through choice literature.

Taking preferences to be partly inferred from the agents' own behaviour gives further reasons for rejecting discovered preference views. If preferences are inferred and affected by previous choices, then learning effects are expected. Similarly, given that agents' preferences are reactive to situational factors in the environment, convergence to stability is to be expected (cf. Hoeffler & Ariely, 1999). An important conclusion follows: if preferences are not revealed or discovered as a result of choice, then they are *constructed* during the choice *process*. Adopting a constructed view, hence, invites us to view preference and choice in a dynamic framework and study their evolution in real time.

Gaze and preferences

> We begin by coveting what we see every day. … And don't your eyes seek out the things you want? – H. Lecter in *The Silence of the Lambs*

Mere-exposure and the gaze-cascade model

The first effects hinting at a role for attention in preference formation came not from studying perceptual processes, but from social psychology and the discovery of the mere-exposure effect (Zajonc, 1968; Moreland & Zajonc, 1976). Zajonc (1968) demonstrated that by simply letting participants view nonsense words and symbols (presented as being in Turkish or Chinese respectively) in varying frequencies it was possible to influence participants' judgments about the positive meaning of those words. The proposed mechanism suggested that exposure altered participants' affective evaluations, and later work showed that this was sufficient to affect the construction of preferences during choice (Baker, 1999).

Building on the mere-exposure effect and finding that infants tend to orient towards novel and preferred stimuli in their environment, Shimojo and colleagues attempted to show that active gaze shapes preferences (Shimojo, Simion, Shimojo & Scheier, 2003). Shimojo and colleagues asked participants to choose between faces and abstract shapes presented side by side in two-alternative forced choice tasks (2-AFC). The critical finding was a marked tendency in participants to look towards the to-be chosen alternative in the final second of viewing leading to their decision (see Fig. 4). This voluntary and gradually increasing self-exposure to the stimuli prior to the decision was called the gaze-cascade *effect*. Shimojo at al. hypothesised that this could be explained by a dual contribution model of preferential decision-making (the gaze-cascade *model*). On this model, choices are an outcome of existing valuations and preferences, but these are, in turn, affected by a feedback loop involving gaze and the environment during the choice process. Agents orient towards their preferred stimulus, but by doing so they also increase their preference, ultimately leading to a choice. In other words, gaze reflects the developing preference and by doing so biases the competition between the options. It might, on first sight, seem that the gaze-cascade model is incompatible with the constructed preference view outlined earlier, since gaze is supposedly driven by preference

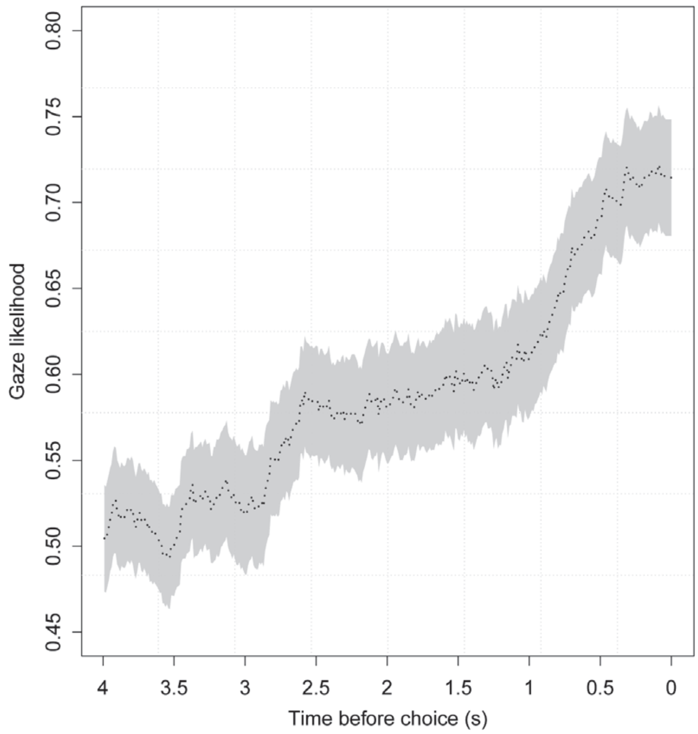

Figure 4. Data from Paper IV illustrating the gaze-cascade effect – an increasing likelihood to gaze towards the to-be chosen option prior to choice.

which is assumed to be constructed during choice. Recall however, that gaze-cascades, which are considered the signature of the feedback loop between preference and gaze, occur only late in the deliberation process during the final seconds leading up to choice. Hence the model is compatible with preferential construction throughout the choice and does not need to assume pre-existing, stable preferences in the agent for gaze to drive the final decision when the gaze cascade is observed.

The gaze cascade effect and the underlying model have been extensively studied and discussed since its original publication. Using the gaze-cascade model of preference formation and a Bayesian prediction network, participants' choices could be accurately predicted in 81% of cases in a neck-tie selection task (Bee, Prendinger, Nakasone, André & Ishizuka, 2006). Gaze-cascades were also found in an 8-AFC using both faces and company logos as stimuli (Glaholt, Wu & Reingold, 2009). The chosen items typically also had the most overall fixation time towards them. Interestingly, the researchers also found evidence for attentional competition between the chosen and the second-most viewed alternatives in the seconds preceding the onset of the gaze cascade. In Paper IV, the existence of gaze-cascades during moral decision-making is demonstrated for the first time, thus considerably extending the types of stimuli and choices for which gaze has been implicated as having an active role during preference construction.

Glaholt and Reingold (2009a) further replicated the gaze-cascade effect, but also found evidence for gaze biases in non-preferential tasks – judging the recency of photographs – suggesting that late orientation towards a to-be chosen option might be a result of a selection process rather than necessarily reflecting preference formation solely. This contradicts the original formulation of the gaze-cascade model, according to which the gaze-cascade is considered a hallmark of preferential decision-making (Shimojo et al., 2003). In Paper III, similar results as Glaholt and Reingold's are reported. There visual, attention towards to-be chosen options was compared during both decision and judgment tasks with varying amounts of visually available information. By analysing both transitions between and within options as well as the time-course of preferential looking, it was possible to discern differences between how visual attention was used to support choice and selection in the different tasks[6]. While visual attention does not necessarily imply preference formation, in a decision task that appears to often be its role.

To provide evidence for the causal claim central to the gaze-cascade model, Shimojo and colleagues introduced a forced exposure paradigm during which participants were shown similarly rated faces sequentially, one at a time, with different exposure times (300ms vs 900ms). One face was presented to the left of the screen and the other to the right. Participants preferred the most exposed face in around 60% of trials. No effect was found when presenting faces centrally on the screen. Since in the latter case no gaze shift was required, this was interpreted as evidence that mere exposure could not account for the effect and that active, preferential looking was the causal mechanism (see also Simion & Shimojo, 2006, 2007; Park, Shimojo & Shimojo, 2010). The forced exposure procedure was replicated for consumer goods by Armel and colleagues (Armel,

[6] Note that the findings in Paper III discern between average gaze patterns between judgments and decisions; they do not speak to the question concerning which varying roles that individual fixations might have during each process.

Beaumel & Rangel, 2008). They found that participants' choices could be manipulated for appetitive options but not for aversive ones. For aversive options more exposure instead decreased the probability of choosing that option.

However, Glaholt and Reingold (2009b) reported failing to find any effects on choice from pre-exposing 4 items in an 8-AFC. They also found larger gaze bias towards chosen items in non-pre-exposed conditions compared to pre-exposed conditions. Both findings were interpreted as evidence against the gaze-cascade model. However, it is not certain if the gaze-cascade model is sufficiently detailed to make specific predictions in such a vastly more complicated setting as Glaholt and Reingold used compared to the 2-AFC situation it was developed in. Additionally, the amount of pre-exposure was comparatively small (1s) making it difficult to firmly interpret the results. Nevertheless, this highlights an important weakness of the gaze-cascade model – its lack of computational specificity.

The aDDM and related models

The findings of Armel et al. (2008) reviewed above, while predicted by the gaze-cascade model, were not interpreted in light of that model but rather in terms of an alternative account for the role of gaze in preference formation; this is the attentional drift-diffusion model (aDDM; Krajbich, Armel & Rangel, 2010; Krajbich & Rangel, 2011; Krajbich, Lu, Camerer & Rangel, 2012). Since it was first proposed, the model has become the dominant approach to understanding the links between eye gaze and choice.

The aDDM assumes that decisions are made as a result of the brain first assigning values to options, and then comparing these values. In the next section, I will return to this value assignment in relation to the constructed preference view that has been discussed earlier. The comparison is assumed to be a diffusion process driven by accumulation of stochastic (relative) evidence integrated to some decision bound (Ratcliff, 1978; Ratcliff & McKoon, 2008; Bogacz, 2007). Originally developed to predict response times in memory retrieval tasks, the diffusion model has been used extensively to study perceptual decision-making. It has been widely found to explain neuronal firing rates in monkeys (Hanes & Schall, 1996; cf. Shadlen & Kiani, 2013) as well as response times in humans (Van Zandt, Colonius & Proctor, 2000). The diffusion model implements a statistically optimal procedure for trade-off between speed and

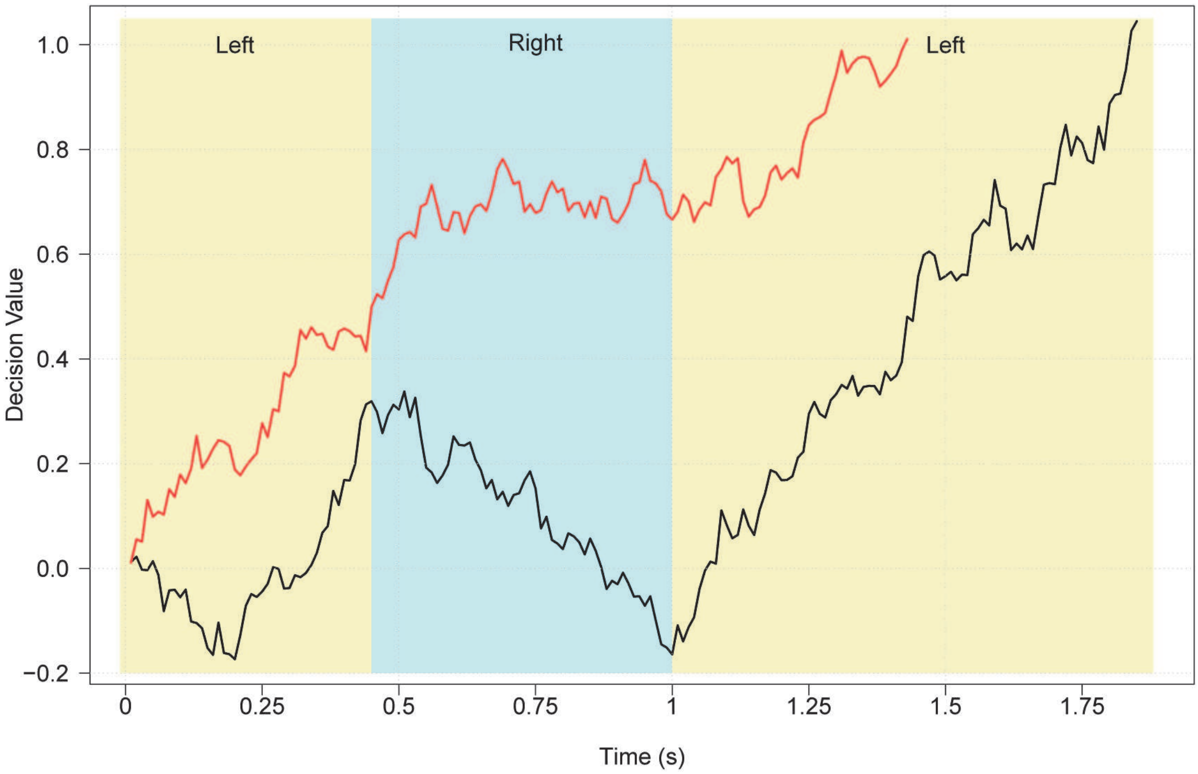

Figure 5. Example of two simulation runs of the aDDM model with varying gaze bias (θ) parameters. Model parameters were set to be d = 0.003, σ = 0.03 and the sample rate to be once every 10ms. The red line represents a model with θ = 0.8, and the black line represents a model with θ = 0.3. The options in both cases had values, left = 6 and right = 5. Gaze direction represented by yellow (left) and blue (right) background.

accuracy with respect to noisy evidence. It has be shown that cortical models, which additionally model neuronal firing rates or competition between neuronal populations (e.g. Usher & McClelland, 2001; Wong & Wang, 2006), can all be reduced to a diffusion model for some set of parameters (Bogasz, Brown, Moehlis, Holmes & Cohen, 2006).

In the Krajbich and Rangel version, the key modification is that the slope of integration – capturing how much evidence is accumulated – varies depending on the direction of the agent's attention. Since attention is assumed to be indexed by gaze, in all applications of the aDDM the slope of integration varies with gaze direction. Assuming the agent starts with a decision value of 0 and decides when reaching either -1 or 1, the model is captured by the following equation (see also, Fig. 5):

$$V_t = V_{t-1} + d\left(r_{fix} - \theta r_{nonfix}\right) + N(0, \sigma)$$

Where V_t is the decision value, d is the parameter controlling the overall drift rate, r is the value of each option, fixated and non-fixated, θ is a gaze bias parameter and N is white Gaussian noise. The theta parameter indicates the magnitude of gaze bias in the decision process, and represents the novel contribution of the aDDM. If theta is one, there is no gaze bias in the model and each time step integrates the full relative value of the options. By contrast, if theta is zero the model exhibits full gaze bias and only the currently fixated options value is integrated.

The aDDM has been shown to be able to account for average response times and average choice frequencies, as a function of the value differences between options, for choices in 2-AFCs between foodstuffs (Krajbich et al., 2010), and consumer durables (Krajbich et al., 2012). An extension to fit trinary choice tasks has also been proposed (Krajbich & Rangel, 2011). It appears that the degree of gaze bias in the decision process, as capture by the theta parameter is highly variable. For foodstuffs, the best fitting model suggested a fairly high degree of gaze bias (θ = 0.3; Krajbich et al., 2010), while for purchasing decisions, a lower degree was found (θ = 0.7; Krajbich et al., 2012). The model can also account for some relations between gaze and choice, for example relations between total gaze time and choice (i.e., more exposure to an option correlates with choosing that option) and that final fixation direction tends to predict choice.

One important aspect of the aDDM model is that it specifies the mechanism by which there can be a causal relationship between gaze and choice. Agents bias their decision process through gazing towards different options, and for equally or similarly valued options this can, in fact, determine the decision. The results forced-exposure experiments discussed earlier have been interpreted as evidence for this (cf. Milosavljevic, Navalpakkam, Koch & Rangel, 2012 for related findings). In Paper V, an alternative method of influencing decisions using gaze was introduced. There, participants' eye gaze was assumed to probabilistically track their trajectory in decision space, based on a continuous cognition view as outlined in the previous chapter. Participants' deliberation was interrupted based on the gaze patterns during the experiment while they made choices about abstract moral principles (see Fig. 1). We found that participants' choices could be influenced in around 58% of trials (cf. Pärnamets et al., 2013 for similar results using non-moral, factual stimuli). When analysing the findings, we found that the most important factor determining choices was final fixation direction compared to relative exposure. This means that some modification might be necessary to the aDDM. In particular, our findings suggest that evidence integration might be leaky, i.e. that past information is discounted compared to present information during the accumulation process (Usher & McClelland, 2001; cf. Koop & Johnson, 2013). This conjecture remains to be tested in future work.

In Papers VI and VII, the aDDM is extended to data for moral decisions. For choices between abstract alternatives, we found similar high levels of gaze bias, $\theta = 0.3$, while for decisions between charitable organisations only slightly lower levels, $\theta = 0.4$. This indicates that a similar diffusion process might be operating even for choices between right and wrong alternatives. However, we found worse fits to the participants' gaze behaviour than previously reported in the literature. This suggests that the fixation process might be different for choices using the materials in Papers VI and VII. Apart from the moral content, one apparent difference is the use of text-based options in Papers VI & VII compared to images of products in previous works.

This last point, concerning how gaze is modelled, raises a key limitation of the aDDM in its current formulation. In the model, fixations drawn from an empirical distribution of durations and transitions between options are modelled as a Markov process; meaning, there is some probability of switching gaze direction after each fixation, but such shifts are imposed exogenously to the decision process. This is in contrast to how gaze-cascades are understood in the gaze-cascade model; there, more preferred alternatives are fixated as a result of the decision process and gaze shifts are thus endogenous on that model. The Markov assumption is unrealistic, because there is a wealth of research indicating how both top-down and bottom-up factors can affect gaze (cf. Orquin & Mueller Loose, 2013). For example, it is known that early on, when first encountering a scene, gaze is strongly driven by low-level saliency features (e.g. visual contrast) of the visual environment (Itti & Koch, 2000). Generalising the saliency approach, it has been suggested that saliency is encoded in the lateral intraparietal area (LIP) of the brain. This area is suggested to code for novelty and reward in the environment, suggesting that visual attention is guided by general information maximisation principles (Gottlieb & Balan, 2010; Gottlieb, Hayhoe, Hikosaka & Rangel, 2014). Recent work has begun to address this by augmenting the aDDM to model gaze as driven both by (low-level) saliency and values in the environment in a 4-AFC task using foodstuffs (Towal, Mormann & Koch, 2013), though much more remains to be learnt about how both features of the environment and task demands shape gaze and the decision process.

Valuation and values

> It is perhaps a testimony to the coerciveness of interview situations how rarely participants say *don't know*, much less try to bolt – B. Fischhoff (1991, p. 841)

In the preceding discussion of the aDDM, I have freely been talking about the values which form the basis for the evidence accumulation on which the diffusion decision framework relies. Recall that the model posits that choice emerges from a comparison process based on the values assigned to the options available to the agent. A natural question arising is how do these values relate to the concept of preferences?

A first step is to consider what goes on during experiments like the ones presented in Papers VI and VII (see also, Krajbich et al., 2010). Participants are first asked to rate the stimulus items that will later become choice options. Following this, a choice phase ensues, consisting of a large number of 2-AFC trials. The model is then fit using both sets of data. The data from all these experiments suggest that these ratings correlate, on average, with what would be expected from a value measurement. That is, participants are faster when choosing between options that are further apart, compared to options rated as being closer [along the relevant value dimension]. Similarly, when choosing between equally valued options, there is about chance probability that either is chosen. As the difference increases between the options, the probability goes to certainty that the agent will choose the higher rated option (cf. Rangel & Clithero, 2012). Why do these ratings correlate with later behaviour and what it is that is being measured?

One seemingly tempting explanation, that I wish to resist, is that ratings correlate with choice because they reveal the underlying preferences of the agent. In a sense, this would amount to resuscitating something like the discovered preference view discussed earlier (Plott, 1996). However, notwithstanding the problems associated with such an account already mentioned, there is already an established alternative to preferences. This alternative is expected reward. In the fields of learning, animal psychology and artificial intelligence agents' systematic and adaptive ability to interact with the environment (e.g. to make decisions about exploration or exploitation) depends on learning from experienced rewards or punishments and associating these with stimuli (cf. Schultz, 2004; Balkenius, 1995; Sutton & Barto, 1998). These associations are stored and later recalled when the stimulus, or one similar to it, is encountered again. What is important to note is that a reward value is not a preference – there are no constraints on learnt values such as completeness or stability. For example, just like memories degrade due to decay for interference (cf. Waugh & Norman, 1965), learnt behaviours are extinguished (Altmann & Gray, 2002), implying that associated values are not temporally stable.

The notion that the rating values assumed by decision models should be homologous with learnt stimulus values is also compatible with recent neurobiological evidence. For example, a great deal of recent research has attempted to locate the areas in the brain that might be responsible for value-based computations. Over a range of tasks, a number of brain areas have been implicated leading to suggestions that the brain implements a common currency for choice (Levy & Glimcher, 2012; Glimcher, 2014). Similar brain regions, orbitofrontal cortex and ventral striatum in particular, have been found to be related to the coding of reward values (Schultz, Dayan & Montague, 1997; Schultz, 2004). Recent findings that brain activity measured by fMRI when viewing stimuli can be used to predict later choice is also compatible with the view that it is values, in the sense of learnt reward or punishment expectations, which are being reflected rather than preference relations (Smith, Bernheim, Camerer & Rangel, 2013).

Returning to an experimental setup like the one described above where participants first rate a number of alternatives and then proceed to make choices between them. In such experiments, participants' choices can thus be understood as reflecting their learnt associations with that stimulus, associations which are also correlated with their ratings. Second, given that they have made prior ratings, participants are also likely remembering these earlier judgments and relating to them, similarly to how measuring purchasing intent can influence later choice (Morwitz et al., 1993). This second suggestion can be tested directly through a false feedback manipulation on ratings, similar to previous choice blindness studies (e.g. Hall et al. 2013).

Finally, accounting for choice as a question of remembering associations bears affinity to a general framework for understanding valuations in choice experiments known as 'decision by sampling' (Stewart, Chater & Brown, 2006). On this theory, decisions are made by local comparisons without the brain calculating anything resembling global values, instead only relative ranks are compared (Ivo, Chater, Stewart & Brown, 2011). These relative ranks are produced by sampling from previous experiences (i.e. from memory) with stimulus of the kind that is chosen between. Decision by sampling, thus, implies a strong role for contexts and for attentional mechanisms which can bias the sampling (cf. Tstetsos, Chater & Usher, 2012).

Most neurobiological theories of decision-making will agree that any value signals in the brain, regardless of source, are normalised based on the decision context, but differ with respect to what kind of comparisons the brain makes during decision-making. These are important questions for future work, both psychological and neuroscientific, but furthering the understanding of how choices evolve over time will not require any strong assumptions of preferences in the decision maker.

Moral cognition

> The very idea of morality implies a force beyond any particular individual, a force that makes demands and punishes transgressions. These demands and punishments are not ordinary ones. You are expected to follow a moral duty, regardless of whether it is useful or injurious to yourself. – R. Collins (1992, p. 38)

Here I take a 'preference and decision' perspective on moral cognition, as is reflected in Papers IV-VII. Before spelling out how those Papers contribute to understanding human morality, a brief review of some aspects of contemporary moral psychology is necessary.

Process of moral judgments

> Another distinction in the study of social attitudes which is sometimes lost sight of is that the cognitive and the affective appraisals may be entirely independent. For example, a group of subjects may agree in their strong dislike of communism. Someone might give them an examination in order to show that the subjects actually do not know what they are talking about. That may very well be true … – L.L. Thurstone (1954, p. 52)

Morality is part of what binds people and societies together (Frazer, 1911; Collins, 1992). However, defining it for the purpose of psychological inquiry is more difficult; to the point where most papers on moral cognition, including Papers IV-VII, do not attempt to give a clear definition of morality. While unambiguous cases of moral behaviour can easily be found – helping an old lady over the street or deciding to have an abortion or not – it seems that generally the concept of morality is a case of family resemblance (Wittgenstein, 1953; Rosch & Mervis, 1973). Similarly, it is clear that morality in some sense is about norms and rules – but not all norms and rules; the offside rule in football is likely not a moral rule, to not murder very likely is, while whether it is acceptable to wipe ones hands on the table cloth might or might not be (cf. Haidt, Koller & Dias, 1993).

For much of the 20th century, empirical moral psychology built on a rationalist conception of morality derived from the philosophers Kant and Bentham (Haidt & Kesebir, 2010; Bartels, Bauman, Cushman, Pizarro & McGraw, 2015). If morality is a function of rational reasoning then the development of morality and reasoning should co-occur. This was the view taken by many influential thinkers, such as Piaget and Kohlberg (e.g. Kohlberg, 1963), and moral psychology was for many years synonymous with developmental psychology.

A second result of adopting the rationalist perspective was that normative ethics became a choice between deontology and consequentialism. Deontology is concerned with duties, and actions are classified as being right or wrong independent of their consequences. Kant's golden rule, stating that actions are right only if one could will the rule governing that action to become a universal maxim, is the most influential method for deriving if one has a duty to perform an action or not. Consequentialism, on the other hand, states that an action is right if it produces desirable consequences. In the version used in most contemporary research, desirable consequences are defined as those that maximise individual welfare. Given the emphasis on acts and welfare, moral psychology concerns itself with the version of consequentialism known as act utilitarianism (Kahane, 2012). The upshot of this was that the stimuli used by moral psychologists primarily concerned actions involving harm to others and of just treatment and rights (Haidt & Kesebir, 2010).

Two methodological developments lead to a break from the rationalist tradition. The first was that researchers in psychology and cognitive neuroscience started to rediscover emotions and affective responses as being an integral part of human cognition (Damasio, 1994; Zajonc, 1980). The second was cross-cultural research which contributed to an understanding of morality as comprised of multiple and varying concerns, dependent on proposed interplays between evolutionary and social factors (Haidt, 2007; Shweder, Mahapatra & Miller, 1987).

The affective revolution in moral psychology can be said to begin with the publication of two papers in 2001. In the first, an fMRI investigation on people responding to high-conflict moral dilemmas showed that areas associated with emotions, such as the angular gyrus, have higher average activation than during

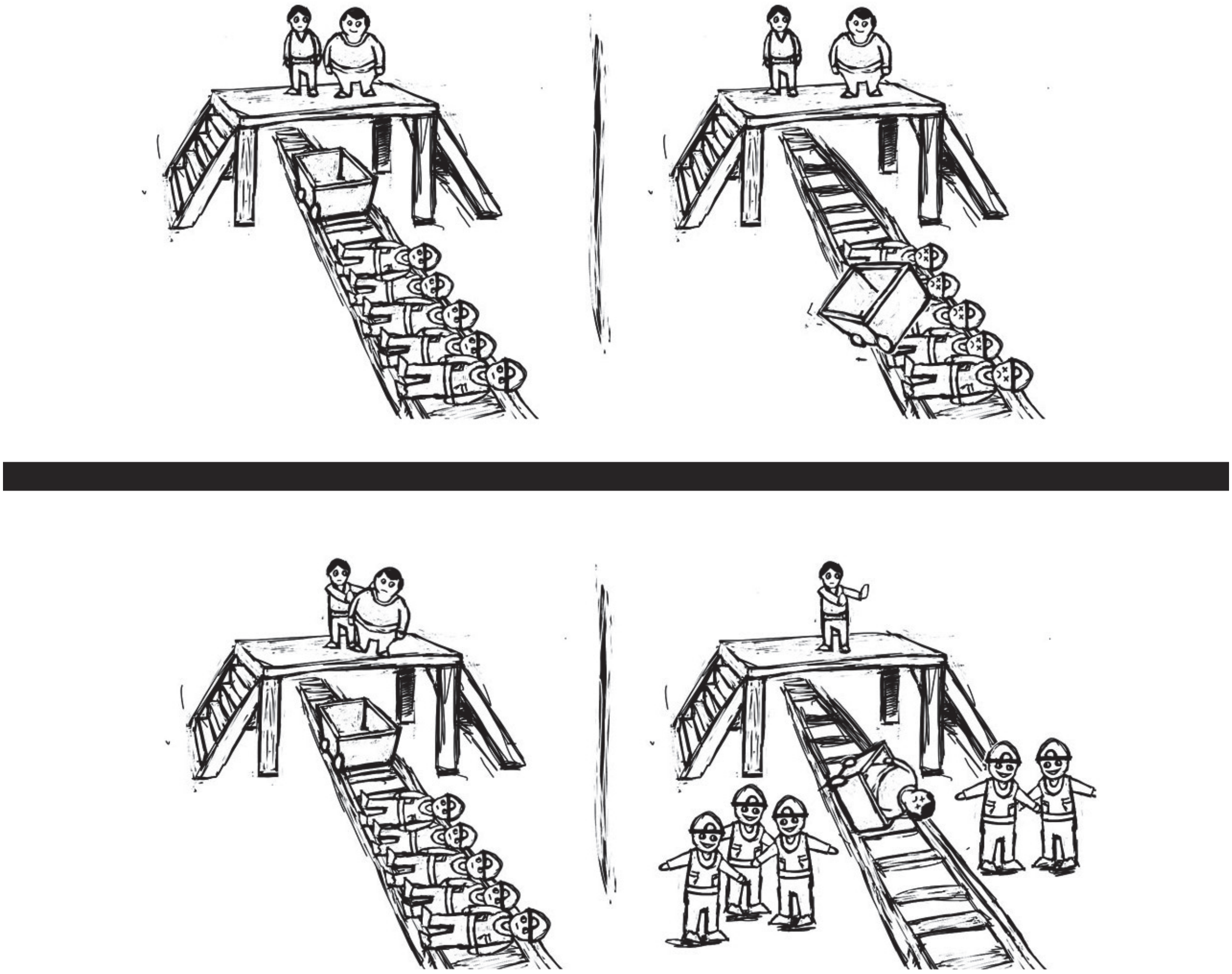

Figure 6. Pictorial example of a Trolley-type moral dilemma (Paper IV), here the 'Footbridge problem' (Thomson,1985). In this problem you are asked to choose one of two options in response to a situation where you are standing next to an obese man on top of a footbridge. Below you there is a trolley racing towards five workmen, who face certain death if the trolley is not stopped. **Top panel.** You do not intervene and the five workmen are killed. This is interpreted as the deontological choice. **Bottom panel.** You push the obese man onto the tracks into the way of the trolley, thereby killing him in the process, but the five workmen survive. This is interpreted as the utlititarian choice.

low-conflict moral dilemmas (Greene, Sommerville, Nystrom, Darley & Cohen, 2001; cf. Moll et al., 2002). The second synthesised a broad range of findings across disciplines related to human behaviour to claim the primacy of intuition over reason in moral judgment (Haidt, 2001). While a consensus had emerged dethroning reasoning as the sole driver of moral choices, the question of the relative role of reasoning remained divisive (Greene & Haidt, 2002).

Two models and their discontents

On one approach, in what has become known as the Dual-Process Model (Greene, 2007; Greene, Morelli, Lowenberg, Nystrom & Cohen, 2008; Paxton & Green, 2010), moral judgments are proposed to arise from the competition between fast, affective responses and slow, deliberate reasoning. This dual-system view mirrors similar distinctions in psychology from reasoning to decision-making (Evans, 2003; Kahneman, 2003). However, Greene's important twist to the dual-system view is the hypothesis that each system, or process, is primarily responsible for one particular kind of moral judgment. This research was pioneered by using a specific form of moral problems called Trolley dilemmas (see Fig 6.), which can be understood as pitting utilitarian versus deontological intuitions against each other (but see Kahane, 2012, for a critique of this assumption). Responses to these dilemmas were then correlated with brain activity, and deontological judgments were mapped to emotionally driven processes while utilitarian judgments were mapped to reason driven processes. For example, studies on patients with damages to the ventromedial prefrontal cortex, a brain area associated with emotional and social evaluation, showed that these patients had an increased frequency of utilitarian judgments (Koenigs et al. 2007). Similarly, increased activations in the anterior cingulate cortex, which is thought to mediate response conflicts, and the dorsolateral prefrontal cortex, associated with cognitive control, were found to be correlated with utilitarian judgments (Greene, Nystrom, Engell, Darley & Cohen, 2004). The reason-based system is assumed to need time to override an initial emotional response. Findings showing that utilitarian judgments require longer response times to make utilitarian judgments and findings that imposing shorter response times decreases utilitarian judgments, have both been interpreted as supporting the dual-process model[7] (Valdesolo & DeSteno, 2006; Suter & Hertwig, 2011).

For Greene, the Dual-Process Model, if correct, has normative implications. If deontological choices are underpinned by emotional processes, but understood by the agent as being rationally chosen, then the agent is objectively deluded about her choices. Furthermore, since deontology is founded on rationalist

---

[7] In Paper IV we do not find this response time pattern, possibly indicating it is less stable than previously hypothesised.

premises, the whole theory as a normative project is proposed to fail (Greene, 2007). However, the proposed strict distinction between processes supporting utilitarian and deontological decisions might not be as straightforward as the Dual-Process Model suggests. Recent work has indicated that higher blood alcohol content, which in turn affects participants' ability of exercising cognitive control, increases the frequency of utilitarian responses to moral dilemmas (Duke & Bègue, 2015). Likewise, participants giving utilitarian responses to moral dilemmas tended to score higher on personality scales indicating Machiavellian, psychopathic and nihilist traits (Bartels & Pizarro, 2011). Similarly, utilitarian response tendencies were found to correlate with egoistic behaviour and less concern for the greater good (Kahane, Everett, Earp, Farias & Savulescu, 2015). What this suggests is that the strong mapping between moral responses and proposed processes underlying moral judgments and choices might be overstated. Instead, the data might be capturing other distinctions, such as that between intuitive and counter-intuitive dimensions (Kahane, 2012), or between model-based and model-free decision systems studied in the reinforcement learning literature (Crockett, 2013; Cushman, 2013).

A second approach, the Social-Intuitionist Model, argues that fast intuitions, which sometimes have an affective base, underlie most moral judgments and choices (Haidt, 2001; Haidt & Björklund, 2008). These intuitions are thought to be partly founded in biologically grounded similarities between all humans (O'Neill & Petrinovich, 1998; Haidt, Rozin, McCauley & Imada, 1997; Haidt & Joseph, 2007), and partly shaped by interacting social and cultural constraints contributing to each individual's morality (Haidt et al. 1993; Graham & Haidt, 2010). Attempts have been made to map the intuitions of the Social-Intuitionist Model to five moral 'foundations' (Graham et al. 2011). This mapping is part of a larger effort to broaden the conception of morality used in empirical moral psychology to one that moves beyond questions of harm and fairness.

In the Social-Intuitionist Model, moral reasoning plays only an indirect part in forming moral judgments; more often, it is considered to be taking a post-hoc, confabulatory role justifying whatever the dominant intuition already has 'decided' (cf. Nisbett & Wilson, 1977; Thurstone, 1954). For example, when probed about reasons for making a moral judgment about cases of a 'harmless' moral transgression, such as sexual acts between consenting adult siblings using contraception, participants often failed to give justifications when their initial reasons were questioned by the experimenter (Haidt, Björklund & Murphy, 2000). Instead, many participants would fall into a pattern of responding with a "just because it is [wrong]", a phenomenon Haidt and colleagues dubbed moral dumbfounding. Similarly, studies have shown that participants fail to provide justifications which can account for their moral judgments in response to more typical moral dilemmas, implying that moral rules or principles that might be construed as governing their moral choices do not seem to be consciously accessible to participants (Hauser, Cushman, Young, Kang-Xing Jin, & Mikhail, 2007; Cushman, Young & Hauser, 2006). Findings of choice blindness for moral (Hall, Johansson & Strandberg, 2012), and political attitudes (Hall et al., 2013), and the subsequent both ability and willingness of participants in those studies to confabulate reasons for the manipulated attitude, further highlights disconnects between moral choices and their justifications. The role given to reason in this theory is thus similar to that proposed in argumentative theories of human reasoning, where reasoning functions as providing arguments to support decisions already made (Mercier & Sperber, 2011).

The Social-Intuitionist Model, despite its reliance on specifically moral intuitions, might be compatible with sets of findings showing that not only are the principles underlying moral judgments inaccessible to people, but they might also be based on non-moral content (Cushman & Young, 2011). For example, when attributing moral faults to animate objects the strength of people's judgments are dependent on kinematic features of those objects. In one study, participants viewed animated videos of cylinders and cones moving and pushing each other so as to cause or prevent the other from falling into a red circle marked as harmful. Participants' judgments were found to depend on whether the objects were moving or stationary, and were also directly related to the speed at which the objects were moving (Illiev, Sachdeva & Medin, 2012; Nagel & Waldmann, 2012). This suggests that moral intuitions, at least concerning blame and responsibility, might be linked to ascriptions of causality (cf. Zultan, Gerstenberg & Lagnado, 2012). Hence, on this alternative account, part of the reason why moral reasoning might be unrelated to the generation moral of intuitions is that there is very little uniquely moral *per se* about how those intuitions are grounded.

Having surveyed these two dominant theories, we can take stock from a more abstract point of view. One important thing to note is that both models, while claiming to capture the process of generating moral judgments and choices, do not actually contain many clear process predictions. This is likely due to how these models are presented and studied; moral cognition is treated as being comprised of a number of cognitive modules, each dedicated towards processing specific forms of information. These modules then discretely combine their output to produce a moral judgment. However, the computational properties of the system are typically not spelled out, and neither is how strictly the modular metaphor is to be interpreted. For the cognitive science of morality to move forward, I believe that specifying process claims is a crucial step that needs to be taken.

A second point is that both models are imprecise concerning how they conceptualise people's preferences. This should maybe not be seen as particularly surprising; the literature on moral cognition is generally shaped by different concerns than that on decision-making, and the question of predicting behaviour from preferences has not been treated with as being important. Nevertheless, there is a long tradition of demonstrating how moral judgments can be manipulated in various was. For example, moral judgments can be made more severe by inducing, through hypnosis, feelings of disgust towards unrelated trigger words used as part of vignettes (Wheatley & Haidt, 2005; see also Zhong & Liljenquist, 2006; Schnall et al., 2008). Similarly, studies have shown that framing can affect moral choices in dilemmatic contexts. Using a save or kill framing[8] altered how acceptable participants judged identical outcomes to be (Petrinovich & O'Neill, 1996; cf. Petrinovich, O'Neill & Jorgensen, 1993; Bartels, 2008). Given this, it seems that any model of moral cognition, for it to be realistic, must adopt some degree of constructed preference approach. The Dual-Process Model, with its emphasis on different, competing processes generating either deontological or utilitarian judgments, seems readily amenable to this view. The intuitions of the Social-Intuitionist Model might, at first glance, seem to be something similar to the standard conception of preferences. However, the model considers intuitions to only be vaguely specified and far from complete, which rules out the traditional view on preferences. Instead the model treats intuitions as a set of core evaluative tendencies in the individual which only receive precise form during the context of eliciting a judgment or making a choice (Haidt & Björklund, 2008). As such, this model also seems compatible with the constructed preference perspective.

---

[8] Note the similarity to Tversky and Kahneman's (1981) gain versus loss framing manipulations, however, with the difference that in Petrinovich's studies participants are not given choices but are asked to rate acceptability.

One possible upshot of this is the following: if moral preferences are constructed like non-moral preferences, might it be possible to study moral choices like other choices are studied? Might even similar processes be found to be underpinning both? With this final query the ground is set to consider the contributions of the Papers of this thesis to moral cognition in large.

A new approach to moral cognition

The papers presented here can be understood as addressing both the process and preference points raised above. The most important contribution is the introduction of a clear time course perspective in the study of how moral choices are formed: to begin to understand how moral decisions unfold over time.

In Paper IV, this is done within the context of the Dual-Process Model, and the tradition of studying morality by using Trolley-type problems. The two main findings are, first, that gaze-cascades are present during moral decision-making, indicating an active role of gaze during moral preference formation (Shimojo et al., 2003). Second, that by examining differences in the distribution of attention and in the dynamics of eye gaze, processing differences between utilitarian and deontological responses was found. The latter finding indicates how eye gaze can be used to study moral cognition in real-time. This suggests that the general framework of embodiment and continuous processing, outlined earlier, also applies to the moral case. Given that framework, finding that eye gaze tracks moral cognition might seem as an expected outcome; nevertheless, demonstrating it puts empirical meat on the theoretical bones.

Furthermore, the methods used in Paper IV, as well as those found in Papers I and III, can easily be used to study moral processes with other stimuli than Trolley-type dilemmas. There is likely a general loop between improved methodological tools and theoretical specificity. Hence, the hope is that as new ways of testing processing claims are developed and used, so will also moral models become increasingly specific about their processing claims.

In Paper V, the embodied processing perspective is applied one step further. There, the hypothesis that where participants are looking is causally connected to what they will end up choosing was tested. The results showed that not only was participants' gaze direction correlated with what they chose, but, by interrupting participants' deliberation based on their eye gaze, their moral choices could be shifted to a randomly predetermined option. It, thus, appears that moral deliberation can be understood as an embodied process, whereby eye gaze continuously tracks a decision maker's trajectory through a moral state space. As such, the timing of our interactions with a task environment, such as options presented on a screen, or our general surroundings, can have a definite impact on what choices we make and what preferences we construct.

If so, like other findings indicating that people accept environmental influences when reasoning about their moral and political attitudes (Hall et al., 2013), one key implication might be that our moral identities are far less stable than common sense suggests. This can be seen as a liberating force. Instead of feeling an obligation to be burdened with the complete, transitive preference ordering required by classical decision theory, perhaps our stance towards moral questions should be reflectively inquisitive and evaluative in the moment of choice.

As such the findings in Paper V are compatible with the Social-Intuitionist Model. Participants were not aware of the gaze-contingent nature of the manipulation and the fact that their choices were being shifted. They even rated manipulated and non-manipulated choices as being equally important. This is further evidence of the post-hoc, confabulatory mechanisms involved in morality. The stimuli used in Paper V targeted the core questions involved in the five dimensions proposed to underpin our moral intuitions (Graham et al., 2011). Hence, the results of Paper V put limits on the representational detail which the intuitions can be proposed to have.

From the perspective of the Dual-Process Model, the findings of Paper V could be understood as targeting the fast, emotional system, given the overall speed of the gaze-contingent manipulation prompt. However, since the stimuli do not have obvious emotional connotations, it is not clear if the model with its sharp distinction between emotional and rational processes can properly accommodate the results.

One interesting question for further study arises when considering the social component in the Social-Intuitionist Model. The model emphasises that agents are socialised into their intuitions (Haidt & Björklund, 2008; cf. Collins, 1992), and that moral reasoning occurs primarily through social interaction. There is clearly no such social component in Paper V. Investigating how moral choices are generated in social contexts is a pressing question for understanding morality outside the lab. Using eye gaze provides a process measure to see if, and how, the choice process might differ in that case.

That eye gaze can be used to measure and influence an agent's moral deliberation also suggests that moral choices can be captured by computational models quantifying how fixations affect the choice process. Papers VI and VII investigate precisely this, and demonstrate that the aDDM model can be fit to data from moral choices. I have already detailed the aDDM previously, so the following discussion will emphasise what these findings entail for understanding moral cognition, and what some of the limitations of the approach might be.

As a starting point, it is worth emphasising that this represents only a first step towards understanding some of the computational properties of moral decision-making, and that the current formulation of the aDDM model is likely an oversimplification of the underlying dynamics. Nevertheless, it is particularly interesting to take a modelling approach in the case of moral decisions, and here two promising aspects of these findings are discussed mentioned.

First, is that they allow for precise predictions of the quantitative state of an agent's moral decision process[9], in a way that has not been anticipated by any hitherto proposed account of moral cognition. This allows for talking about decision-making across different tasks and fields using the same language. For example, the fitted models presented in Papers VI-VII consistently indicate that the exact relationship between eye gaze and choice, specifically with regard to predictions of specific fixation related biases, might differ in the moral case compared to what has been found in other applications of the aDDM (e.g. Krajbich et al., 2010). Even if it turns out that moral and non-moral decisions might require alternative parametrisations of the same model, this work highlights the possibility for a domain general account of both moral and non-moral aspects of human decision-making

A second point concerns how developing a computational understanding of moral decisions might tie in with other developments in the study of moral cognition. Recent work suggests that models of reinforcement learning, long used to understand how values in the environment are learnt and acted upon (Sutton & Barto, 1998), might be applicable to studying moral decision-making

[9] Recall that the aDDM computes a moment-by-moment decision value, meaning that for each moment of deliberation there is, in principle, a prediction of what the currently most favoured option might be.

(Crockett, 2013; Cushman, 2013; Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014). In decision neuroscience, one current major concern is how to connect the study of value learning with that of decision-making. In other words, how to bridge the gap between the aspects of decisions modelled using a reinforcement learning framework with those modelled using a diffusion decision framework (cf. Glimcher, 2014). The tantalising suggestion here is that a similar synthesis between learning and diffusion framework might be forthcoming also for moral decision-making, and perhaps sooner than anticipated.

Viewing the aDDM model through the lens of morality highlights both the power and weakness of the general approach, as a framework of understanding decision-making. By common sense, it seems implausible that a modelling framework developed to capture how the brain responds to simple perceptual stimuli, like random dot motion, also, without any specific modifications, should capture how we choose between moral options. There is great explanatory strength in the simplicity of a model like the aDDM, but in this simplicity a lot of things that might be relevant to moral choice are treated very abstractly. At the same time, it is instructive to compare general moral models with simple diffusion framework. How should the concerns about emotions, intuitions and the social context be understood? The full answers to these questions are for future work to discover, but by taking the first steps, Papers VI-VII show that these are very much empirical questions – from now on, what remains is a modelling choice in how much detail one wants to capture when trying to understand moral decisions!

Looking forward

> The Road goes ever on and on
> Down from the door where it began.
> – B. Baggins in *Lord of the Rings*

The findings in this thesis make three primary contributions to cognitive science and our understanding of human cognition. First, the work here provides fresh evidence for the utility of adopting an embodied and continuous view of cognition in decision-making research. In future work, one particularly interesting prospect is the combination of multiple sources of information, in addition to using eye gaze, to capture different aspects of the choice process. For example, eye-tracking can be combined with mouse-tracking (cf. Koop & Johnson, 2013), but also with a number of other techniques such as GSR, EEG and EMG. Motion capture might also provide an additional window onto developing cognition, and combining this with virtual reality displays can allow for the study of decisions under more realistic settings or with scripted interaction. This is important for several reasons. Doing so can give more data on the timing of decisions and provide richer input to models of higher cognition and its sensorimotor interactions. This will allow for the development of more fine-grained hypothesis and, in time, new experimental paradigms allowing us to ask new questions about how the cognitive system works. More data might also help shine light on one important outstanding question, partly touched on in the findings of Paper V, namely the role of sensorimotor feedback for cognitive processes. It is generally the case that knowing how someone is moving, swaying or otherwise reacting, as part of thinking or deciding, can allow one to bias those processes? Does the way I move my arms and eyes shape my thinking, in the same way my thinking shapes those movements?

Second, by taking the idea that preferences unfold over time seriously, the findings in this thesis provide novel support for a constructed preference view. In two of the most remarkable findings, those of Paper II and Paper V, choices are shown to be dependent on beliefs about past choices and current gaze direction, respectively. Together they suggest that preferences can be shaped by powerful forces in the moment as well as by beliefs about past choices. An important question for future work is to understand how to merge these two influences on choice. One route, for example, might be to study the interactions between participants in group-decision studies, where, the public nature of communication can be exploited to try to measure exactly how and where beliefs enter into the choice process. It is also important to understand how factors such as reasons and beliefs enter into the decision process in individuals. Likewise to begin to study how reasons and beliefs can be understood within existing modelling frameworks which operate on values and eye gaze. A starting point for this might be to investigate if agents are sensitive to particular reasons or evidence at specific times during the decision process, for example, when is it best to time an offer of a price reduction during a consumer decision?

Third, this thesis introduces the use of eye gaze to study moral cognition. The key idea motivating this work has been to study moral decision processes largely unburdened by many of the considerations found in regular research on moral choice. The upshot of this has been that the work here has begun to develop an understanding how mechanisms based on evidence integration to a threshold, similar to ones found for non-moral decisions, might underlie moral choices. In the preceding chapter I outlined some ways to move forward with this project, and in particular, to integrate it with work on value learning. Nevertheless, at some point the particulars (real or perceived) of moral psychological content should be taken into account and melded with the approach taken in this thesis. One promising avenue might be to investigate the time course of decisions ascribing blame or praise to agents and their actions, both in terms of relative judgments – "who of these is most blameworthy" – as well as absolute judgments – "is this blameworthy". This could be expanded by borrowing from the seminal work of Heider and Simmel (1944), to attempt to understand how visual attention and interaction dynamics blend to form ascriptions of blame and praise in agents.

Another prospect for future work in the moral domain is to use the methods developed in Paper II. After all, many moral questions are based on communal decisions. Using a choice blindness method to dissect how group members construct their moral decisions might yield new insights to how this process evolves, for example in relation to later conformity with agreed upon rules and norms. Such an approach could also be used to investigate cooperation in game-like settings, an area which has recently received increased attention (Rand, Greene & Nowak, 2012).

To conclude these introductory chapters, my hope is that the work here can serve as a starting point in a new approach towards understanding human morality as well as generate new insights and hypotheses about how human choice and preference construction work in general.

References

Aks, D. J., Zelinsky, G. J., & Sprott, J. C. (2002). Memory across eye-movements: 1/f dynamic in visual search. *Nonlinear Dynamics, Psychology, and Life Sciences*, *6*(1), 1-25.

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, *137*(2), 208-216.

Ariely, D., & Norton, M. I. (2008). How actions create–not just reveal–preferences. *Trends in Cognitive Sciences*, *12*(1), 13-16.

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 396-403.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89-195.

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus

Baker, W. E. (1999). When can affective conditioning and mere exposure directly influence brand choice? *Journal of Advertising*, *28*(4), 31-46.

Ballard, D. H. (1991). Animate vision. *Artificial intelligence*, *48*(1), 57-86.

Balkenius, C. (1995). *Natural intelligence in artificial creatures*. Lund University Cognitive Studies, 37.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*(2), 381-417.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*(1), 154-161.

Bartels, D.M., Bauman, C.W., Cushman, F.A., Pizarro, D.A., & McGraw, A.P. (2015). Moral Judgment and Decision Making. In G. Keren, & G. Wu (Eds.) *The Wiley Blackwell Handbook of Judgment and Decision Making.* Chichester, UK: Wiley.

Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience*, *18*(7), 2047-2058.

Bee, N., Prendinger, H., Nakasone, A., André, E., & Ishizuka, M. (2006). Autoselect: What you want is what you get: Real-time processing of visual attention and affect. In E. André, L. Dybkjær, W. Minker, H. Neumann & M. Weber (Eds.) *Perception and Interactive Technologies* (pp. 40-52). Berlin Heidelberg: Springer.

Bem, D. J. (1965). An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology*, 1(3), 199-218.

Bem, D. J. (1972). Self-perception theory. *Advances in experimental social psychology*, *6*, 1-62.

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, *11*(3), 118-125.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700-765.

Braga, J., & Starmer, C. (2005). Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics*, *32*(1), 55-89.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, *52*(3), 384.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, *47*(1), 139-159.

Chen, M. K., & Risen, J. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. *Journal of Personality and Social Psychology*, *99*(4), 573-594.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*(5), 801-814.

Collins, R. (1992) *Sociological Insight: An Introduction to Non-Obvious Sociology 2nd ed.* New York, NY: Oxford University Press.

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*(6), 1052-1075.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological Science*, *17*(12), 1082-1089.

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363-366.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, *111*(48), 17320-17325.

Cushman, F. (2013) Action, outcome, and value: a dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273-292.

Dale, R. (2008). The possibility of a pluralist cognitive science. *Journal of Experimental and Theoretical Artificial Intelligence*, *20*(3), 155-179.

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York, NY: Putnam.

Delamater, A. R. (2004). Experimental extinction in Pavlovian conditioning: behavioural and neuroscience perspectives. *Quarterly Journal of Experimental Psychology Section B*, *57*(2), 97-132.

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition*, *134*, 121-127.

Duran, N. D., & Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas in Psychology*, *32*, 131-146.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380-417.

Egan, L. C., Bloom, P., & Santos, L. R. (2010). Choice-induced preferences in the absence of choice: Evidence from a blind two choice paradigm with young children and capuchin monkeys. *Journal of Experimental Social Psychology*, *46*(1), 204-207.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*(2-3), 195-225.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454-459.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*(2), 203-210.

Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition*, *119*(2), 179-196.

Fischhoff, B. (1991). Value elicitation: is there anything in there? *American Psychologist*, *46*(8), 835-847.

Frazer, J.G. (1911) *The Golden Bough: A Study in Magic and Religion*. London, UK: Macmillian.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*(4), 673-690.

Gazzaniga, M.S. (2010). Neuroscience and the correct level of explanation for understanding mind. An extraterrestrial roams through some neuroscience laboratories and concludes earthlings are not grasping how best to understand the mind–brain interface *Trends in Cognitive Sciences, 14*, 291-292.

Glaholt, M. G., & Reingold, E. M. (2009a). The time course of gaze bias in visual decision tasks. *Visual Cognition*, *17*(8), 1228-1243.

Glaholt, M. G., & Reingold, E. M. (2009b). Stimulus exposure and gaze bias: A further test of the gaze cascade model. *Attention, Perception, & Psychophysics*, *71*(3), 445-450.

Glaholt, M. G., Wu, M. C., & Reingold, E. M. (2009). Predicting preference from fixations. *PsychNology Journal*, *7*(2), 141-158.

Gibson, J.J. (1986). *The Ecological Approach to Visual Perception*. New York, NY: Psychology Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650-669.

Glimcher, P.W. (2010). *Foundations of Neuroeconomic Analysis*. New York, NY: Oxford University Press.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390-394.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*(1), 10-16.

Gottlieb, J., & Balan, P. (2010). Attention as a decision in information space. *Trends in Cognitive Sciences*, *14*(6), 240-248.

Gottlieb, J., Hayhoe, M., Hikosaka, O., & Rangel, A. (2014). Attention, Reward, and Information Seeking. *The Journal of Neuroscience*, *34*(46), 15497-15504.

Graham, J., & Haidt, J. (2010). Beyond beliefs: Religions bind individuals into moral communities. *Personality and Social Psychology Review*, *14*(1), 140-150.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. Journal of Personality and Social Psychology, *101*(2), 366-385.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving guiding attention guides thought. *Psychological Science*, 14(5), 462-466.

Greene, J.D., & Haidt, J. (2002). How (and where) does moral judgment work?. *Trends in Cognitive Sciences*, *6*(12), 517-523.

Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.) *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (pp. 35-79). Cambridge, MA: MIT Press.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144-1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105-2108.

Grossberg, S. (1980a). Biological Competition: Decision Rules, Pattern Formation, and Oscillations. *Proceedings of the National Academy of Sciences , 77*(4), 2338-2342.

Grossberg, S. (1980b). How does a brain build a cognitive code?. *Psychological Review*, *87*(1), 1-49.

Grüne, T. (2004). The problems of testing preference axioms with revealed preference theory. *Analyse & Kritik*, *26*(2), 382-397.

Gärdenfors, P. (2008). Cognitive Science: from Computers to Ant Hills as Models of Human Thought. In P. Gärdenfors, & A. Wallin (Eds.), *A Smorgasbord of Cognitive Science*. (pp. 11–34). Nora, Sweden: Nya Doxa.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382-385.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*(5827), 998-1002.

Haidt, J., & Björklund, F. (2008). Social intuitionists answer six questions about moral psychology. W. Sinnott-Armstrong (Ed.), *Moral psychology (Vol 2). The cognitive* *science of morality: Intuition and diversity* (pp. 181-217). Cambridge, MA: The MIT Press.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3 pp. 367–391). New York: Oxford.

Haidt, J., & Kesebir, S. (2010). Morality. In S.T. Fiske, S.T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (pp. 797-832). Hoboken, NJ: Wiley.

Haidt, J., Björklund, F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason*. Unpublished manuscript, University of Virginia.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. *Journal of Personality and Social Psychology*, *65*(4), 613-628.

Haidt, J., Rozin, P., McCauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology & Developing Societies, 9*(1), 107-131.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347-356.

Hall, L., & Johansson, P. (2008). Using choice blindness to study decision making and introspection. In P. Gärdenfors, & A. Wallin (Eds.), *A Smorgasbord of Cognitive Science*. (pp. 267-83). Nora, Sweden: Nya Doxa.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE*, *7*(9), 1-8. e45457.

Hall, L., Johansson, P., Tärning, B., Sikström, S., & Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*, *117*(1), 54-61.

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B. & Johansson, P. (2013). How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions. *PLoS ONE*, *8*(4), 1-6. e60554.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427-430.

Harmon-Jones, E., & Harmon-Jones, C. (2002). Testing the action-based model of cognitive dissonance: The effect of action orientation on postdecisional attitudes. *Personality and Social Psychology Bulletin*, *28*(6), 711-723.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335-346.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, *22*(1), 1-21.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243-259.

Hoeffler, S., & Ariely, D. (1999). Constructing stable preferences: A look into dimensions of experience and their impact on preference stability. *Journal of Consumer Psychology*, *8*(2), 113-139.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554-2558.

Iliev, R. I., Sachdeva, S., & Medin, D. L. (2012). Moral kinematics: The role of physical factors in moral judgments. *Memory & Cognition*, *40*(8), 1387-1401.

Johansson, P., Hall, L., & Sikström, S. (2008). From change blindness to choice blindness. *Psychologia*, *51*(2), 142-155.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. *Science*, *310*, 116-119.

Johansson, P., Hall, L., Sikström, S., Tärning, B. & Lind, A. (2006). How something can be said about Telling More Than We Can Know. *Consciousness and Cognition*, *15*, 673-692.

Johansson, P., Hall, L., Tärning, B., Sikström, S., & Chater, N. (2014). Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it! *Journal of Behavioral Decision Making*. *27*(3), 281–289.

Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*(1), 236-242.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441-480.

Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind & Language*, *27*(5), 519-545.

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, *134*, 193-209.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), 697-720.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133-156.

Kelso, J.A.S. (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, *28*(12), 3017-3029.

Kohlberg, L. (1963). The development of children's orientations toward a moral order. *Human Development*, *6*(1-2), 11-33.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908-911.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151-185.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852-13857.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292-1298.

Krajbich, I., Lu, D., Camerer, C., and Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*:193. doi: 10.3389/fpsyg.2012.00193

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, *35*(6), 1009-1051.

Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *383*, 74-77.

Lakoff, G. & Johnson. M. (1980) *Metaphors We Live By.* Chicago: University of Chicago Press.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027-1038.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46-55.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' Acceptance of Real-Time Speech Exchange Indicates That We Use Auditory Feedback to Specify the Meaning of What We Say. *Psychological Science*, *25*(6), 1198-1205.

Lucretius (2005) *Sensation and Sex*. London, UK: Penguin.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *The Bell Journal of Economics*, 587-608.

Marr, D. (1982) *Vision*. Cambridge, MA: MIT Press.

McLaughlin, O., & Somerville, J. (2013). Choice blindness in financial decision making. *Judgment and Decision Making*, *8*(5), 561-572.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(02), 57-74.

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22-24.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141-144.

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, *22*(1), 67-74.

Minsky, M. (1974). A framework for representing knowledge. *MIT Lab Memo No. 306*

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience*, *22*(7), 2730-2736.

Moreland, R. L., & Zajonc, R. B. (1976). A strong test of exposure effects. *Journal of Experimental Social Psychology*, *12*(2), 170-179.

Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does measuring intent change behavior? *Journal of Consumer Research*, *20*, 46-61.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387-391.

Nagel, J., & Waldmann, M. R. (2012). Force dynamics as a basis for moral intuitions. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 785-790). Austin, TX: Cognitive Science Society.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113-126.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939-973.

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: a review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190-206.

Park, J., Shimojo, E., & Shimojo, S. (2010). Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proceedings of the National Academy of Sciences*, *107*(33), 14552-14555.

Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, *2*(3), 511-527.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*(1), 87-131.

Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring Constructed Preferences: Towards a Building Code. *Journal of Risk and Uncertainty*, *19*(1), 243-270.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*(3), 145-171.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*(3), 467.

Plott, C. R. (1996). 'Rational individual behavior in markets and social choice processes: the discovered preference hypothesis', in K. Arrow, E. Colombatto, M. Perelman and C. Schmidt, eds., *Rational Foundations of Economic Behavior* (pp. 225-250). London: McMillian and St, Martin's.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160-174.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*(1), 381-410.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M.J., & Richardson, D.C. (2013). Changing minds by tracking eyes: Dynamical systems, gaze and moral decisions. In M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3127-3132). Austin, TX: Cognitive Science Society.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427-430.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873-922.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368-373.

Rogers, T. T., & McClelland, J. L. (2014). Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, *38*(6), 1024-1077.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533-536.

Van Gelder, T., & Port, R. F. (1995). 'It's about time: An overview of the dynamical approach to cognition', R.F. Port, & T. Van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*, (pp. 1- 43). Cambridge, MA: MIT Press.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 61-71.

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, 243-253.

Sauerland, M., Sagana, A., & Otgaar, H. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology*, *18*(2), 371-381.

Sagana, A., Sauerland, M., & Merckelbach, H. (2013). Witnesses' blindness for their own facial recognition decisions: A field study. *Behavioral Sciences & the Law*, *31*(5), 624-636.

Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, *19*(12), 1219-1222.

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology*, *14*(2), 139-147.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.

Schöner, G. (2008). Dynamical systems approaches to cognition. In R. Sun (Ed.) *Cambridge handbook of computational cognitive modelling* (pp. 101-126). Cambrdige, UK: Cambridge University Press.

Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do decisions shape preference? Evidence from blind choice. *Psychological Science*, *21*(9), 1231-1235.

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*(3), 791-806.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317-1322.

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan, & S. Lamb (Eds.) *The Emergence of Morality in Young Children* (pp. 1-83). Chigaco, IL: University of Chicago Press.

Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision research*, *46*(20), 3331-3335.

Simion, C., & Shimojo, S. (2007). Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, *69*(4), 591-595.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129-138.

Skinner, B. F. (1948). 'Superstition'in the pigeon. *Journal of Experimental Psychology*, *38*(2), 168-172.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *The American Economic Review*, 596-605.

Steenfeldt-Kristensen, C., & Thornton, I. M. (2013). Haptic choice blindness. *i-Perception*, *4*(3), 207.

Smith, A., Bernheim, B. D., Camerer, C., & Rangel, A. (2013). *Neural activity reveals preferences without choices* (No. w19270). National Bureau of Economic Research.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review*, *25*(3), 278-298.

Spivey, M.J. (2007). *The Continuity of Mind.* New York, NY: Oxford University Press.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393-10398.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276-315.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1-26.

Strogatz, S. (1994) *Nonlinear Dynamics and Chaos*. Cambridge, MA: Perseus.

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454-458.

Sutton, R.S. & Barto, A.G. (1998) *Reinforcement Learning.* Cambridge, MA: MIT Press.

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Organizational behavior and human decision processes*, *65*(3), 252-267.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Taya, F., Gupta, S., Farber, I., & Mullette-Gilman, O. A. (2014). Manipulation Detection and Preference Alterations in a Choice Blindness Paradigm. *PloS ONE*, *9*(9), e108515.

Thomson, J. J. (1985). Double effect, triple effect and the trolley problem: Squaring the circle in looping cases. *Yale Law Journal*, *94*(6), 1395-1415.

Thurstone, L. L. (1954). The measurement of values. *Psychological Review*, *61*(1), 47-58.

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, *110*(40), E3858-E3867.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1), 86-94.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*(24), 9659-9664.

Tuller, B., Case, P., Ding, M., & Kelso, J. A. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 3-16.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550-592.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476-477.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, *7*(2), 208-256.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*(2), 89-104.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780-784.

White, P. (1980). Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bem. *Psychological Review*, *87*(1), 105-112.

Wilson, A.D., & Golonka, S. (2013) Embodied cognition is not what you think it is. *Frontiers in Psychology 4*(58). doi: 10.3389/fpsyg.2013.00058

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625-636.

Wittgenstein, L. (1953). *Philosophical Investigations*. Chichester, UK: Wiley.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, *1*(6), 209-216.

Wong, K. F., & Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience*, *26*(4), 1314-1328.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2p2), 1-27.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151-175.

Zelinsky, G. J., Rao, R. P., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, *8*(6), 448-453.

Zhong, C-B. & Liljenquist, K. (2006) Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*, 1451-1452.

Zultan, R. I., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Causality and counterfactuals in group attributions. *Cognition*, *125*(3), 429-440.