Knowing you, changing me: Individual differences in social learning and influence in discounting tasks

---

Louisa Thomas

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology

Royal Holloway, University of London

2021

For Rio and Jago

Abstract

Our decisions are often made within social contexts, and our behaviour and beliefs are subject to change due to the influence of other individuals or groups. Contagion is an implicit social influence effect whereby learning about others makes us more like them, and contagion effects have been found in discounting tasks. However, little research has explored individual differences in this effect. The key aim of this thesis is to investigate these differences. Across four experimental chapters, a Bayesian neurocomputational approach was used to examine contagion, and each chapter focuses on different aspects of individual differences in this effect. In chapter 3, contagion of temporal discounting preferences was explored in two neurotypical samples, and one sample of autistic adults. In chapter 4, similarities and differences between contagion of temporal and probability discounting preferences were explored. In chapter 5, the effect of social distance on contagion was explored in an fMRI study. In chapter 6, an online study was conducted to explore the relationship between temporal discounting contagion, and advice-taking and choice behaviour in a social probabilistic reward learning task. In all chapters, a significant contagion effect was found, indicating that contagion is a strong and reproducible effect. However, none of the studies presented within this thesis were able to index individual differences in contagion. No group differences were found between neurotypical and autistic participants, and no significant relationship was observed between autistic traits and contagion in neurotypical samples. A significant relationship between contagion and accuracy was also not found, and contagion also did not significantly differ dependent on the identity of the other agent, despite finding an effect of social distance on learning and feedback processing. In the general discussion, further ideas for exploring individual differences in this effect are explored.

Acknowledgements

Firstly, thank you to Josh, I feel so lucky to have had you as my supervisor. I have learned so much from you, and your feedback has improved my work greatly. Thank you for sitting through all the first draft presentations, for all the post-meeting chats, and (very importantly) for all the cat pictures.

Thank you to Ryan for stepping in and supporting me and cheering me on, and thank you to Dawn, Narender, Jonas and Cat for your support too. Thank you to all the staff at Royal Holloway that made my time here so enjoyable, that I came back for more after my BSc.

Thank you also to all of the PS2010 teaching team that I've had the pleasure of working with over these past years, and to all the students I have had the pleasure of teaching. I have loved every minute (yes, even the marking!).

Thank you to my collaborators Mona and Pat. Your feedback on my work has been invaluable. Pat, thank you also for your support, and for being a friend.

Thank you to Grace for being there for me through both my BSc and my PhD (and Sahira, Callum, Jess, Liam and Tom for your friendship and support, especially during lockdown!). Thank you to Franzi for being a wonderful bouldering buddy and friend (and gin buddy along with Aysha and Arno!). Thank you, Jas and Clare, too for your support throughout (especially writing time!). Thank you also to Becky and Bea for being such wonderful desk buddies! Thank you to **all** of the wonderful people I have had the pleasure of completing my PhD alongside. I am so lucky that I got to spend my PhD with such wonderful PhD pals in the office, many of whom will be friends for life. To all of you. Thank you.

Thank you to my best friend Kerry, for just being there always, for listening to me vent, during my PhD, and all the years leading up to it (and also beyond!). Thank you for always believing in me. You are amazing.

Thank you to my parents, my brother, and to Mark. Your support means the world. Thank you to Mark for keeping me sane (mostly) and supporting me whilst I finished my thesis during the pandemic. Thank you also for making sure I ate and slept, and had enough tea during thesis writing!

Finally (although they won't read it!) thank you to Susie, Jimmy, Jeff, Baloo, Rolo, and Paul, for all the cuddles and love.

Table of contents

Table of tables

Table of figures

sample). At Self2, the posterior has shifted towards the posterior for Other1 (i.e., contagion).		
Figure 3. Group comparisons of D_{KL} scores: Descriptive statistics for all divergence score (D_{KL}) variables for the two NT samples (Study 1 and Study 2), and the ASC sample (Study 3) are presented here, split for more impulsive (log k +1) and more patient (log k -1) agents. Positive values indicate a shift towards the TD preferences of the other agent (i.e., contagion), whereas negative values indicate a shift away. One-sample t-tests (or Wilcoxon signed-rank tests where there were non-normal distributions) were conducted on Study 1 and Study 2 data, and significant results ($p < .05$) are indicated with an asterisk.		**20**
Figure 4 Correlations between percentage accuracy and D_{KL}. A) Shows the correlation for more impulsive agents in the Study 1 sample, and B) for more patient agents in the Study 1 sample. C) Shows the correlation for impulsive agents in the Study 2 sample, only this correlation was significant. D) Shows the correlation for patient agents in the Study 2 sample.		**22**
Figure 5 Histograms and density plots with comparisons between the NT subset and ASC sample. A) Shows the comparison between samples for percentage accuracy for the impulsive agent, B) shows the comparison between samples for percentage accuracy for the patient agent, C) shows the comparison between samples for D_{KL} for the impulsive agent, and D) shows the comparison between samples for D_{KL} for the patient agent.		**27**
Supplemental Figure 1. Descriptive statistics for all shift variables for the two neurotypical samples (Study 1 and Study 2) are presented here, split for more impulsive (i.e., Imp., positive, log k +1) and more negative (i.e., Pat., negative, log k -1) agents. A positive value indicates a shift in value preference towards the other agent (for example, becoming more impulsive after being exposed to an impulsive other) and a negative value indicates a shift in behaviour away from the other agent (for example, becoming more patient after learning the preferences of a more impulsive other). A) Shows normalised basic shift variables, and B) shows normalised distance-controlled variables.		**46**
Chapter 4 (starts on page 54 of thesis)		
Figure 1. Experimental design. A) Shows the order and type of the five blocks for each task. The Self, Other1 and Other2 colours were		**8**

randomised across participants (RGB colours). B) Shows an example experimental screen for Self-trials on the TD task, participants see the type of trial (Self or name of Other), indicated by the name in colour, along with the options. The yellow box appears around the participant's selected choice, following key press. C) Shows an example experimental trial for Other-trials on the TD task, which follow the same format as Self-trials, with additional feedback. This feedback is presented on screen 500 ms after the yellow box appears, and is displayed for 1000 ms before the screen clears for the next trial. D) Shows an example experimental screen for a Self-trial in the PD task, which follows the same format as the TD task. E) Shows an example experimental screen for an Other-trial in the PD task, which also follows the same format as the TD task.		
Figure 2. Correlation between contagion across tasks. This shows the correlation between contagion (as measured by normalised D_{KL}) for the negative (i.e., patient and risk-seeking) others on both discounting tasks. Positive values indicate a shift towards the preferences of the other, and negative values indicate a shift away.		**21**
Figure 3. Task and direction comparisons of normalised D_{KL} scores: Descriptive statistics for all divergence score (D_{KL}) variables for the two discounting tasks, split for direction (i.e., positive, impulsive/risk-aversive and negative, patient/risk-seeking). Positive values indicate a change towards preferences of the other agent (i.e., contagion), whereas negative values indicate a shift away.		**22**
Figure 4. Task and direction comparisons of accuracy: Descriptive statistics for all accuracy variables (accuracy scored out of 30) for the TD and PD tasks, split for direction (i.e., positive, impulsive/risk-aversive and negative, patient/risk-seeking).		**25**
Chapter 5 (starts on page 55) of thesis		
Figure 1. Participant relationship with Close Other. This shows the length of relationship with the close Other for all participants.		**6**
Figure 2. Training task experimental design. This shows the three experimental screens (Offer, Choice and Feedback stages) shown to participants in the training task and the timing of a single trial.		**9**
Figure 3. Task order and experimental screens. This shows the order and type of the five blocks for each task. The Self, Other1 and Other2 text colours were randomised across participants (RGB colours). In all Self-trials, participants saw the Offer screen (options presented on top of each		**10**

other), and the Choice screen (options side by side). In all Other-trials, participants saw the Choice and Offer screens, as well as an additional feedback screen.		
Figure 4. Regions of interest. This shows the seven ROIs used in this paper, identified by Garvert et al (2015). A), B), and C) are presented in yellow, and show the three ROIs associated with contagion in the Garvert paper, and D), E), F), and G) are presented in red, and show the four ROIs associated with learning in the Garvert paper.		**20**
Figure 5. Comparison of D_{KL} for Close and Distant Others. Descriptive statistics for divergence score (D_{KL}) variables, split by identity of the Other (i.e., Close/Distant).		**22**
Figure 6. Comparison of accuracy for Close and Distant Others. Descriptive statistics for participant accuracy (out of 40) at making choices on behalf of the other agent, split by identity of the Other (i.e., Close/Distant).		**22**
Figure 7. Comparison of RTs for each agent. Descriptive statistics for RT variables, for Self (averaged across blocks), Close other, and Distant Other.		**23**
Figure 8. Left postcentral gyrus. This shows the significant Cluster of 370 voxels with greater activation for Close versus Distant Others at the feedback stage, identified in the whole brain analysis (peak [-46, -24, 61]).		**24**
Supplemental Figure 1. Comparison of RTs for each agent. Descriptive statistics for RT variables, for Self1, Self2, Self3, Close other, and Distant Other.		**40**
Chapter 6 (starts on page 56) of thesis		
Figure 1. Temporal discounting contagion task experimental screen. This shows an example experimental screen for an Other-trial. Participants saw the two options, along with a question at the top, indicating who they were choosing for (i.e., *"What option would you choose for this other person?"* for Other-trials, or *"Which option would you prefer?"* for self-trials). The items remained on the screen until the participant had selected their response, before a tick or cross (i.e., feedback) appeared in between the two choice options. Following the feedback, the next trial appeared immediately.		**8**
Figure 2. Card game experimental screen. A) Shows an example experimental screen with reward magnitudes displayed on each card, the recommended card displayed in the bottom right-hand corner, and the score (i.e., number of points collected throughout the task) in the top		**13**

negative/patient) in **chapter 3** (TD task). B) Shows percentage accuracy for the sample presented in **chapter 6** (TD task). C) Shows percentage accuracy for the TD and PD tasks in **chapter 4**, split by direction (i.e., positive: impulsive/risk-aversive and negative: patient/risk-seeking). D) Shows percentage accuracy in the TD task presented in **chapter 5**, split for Close and Distant Others.		

Chapter 1.

General Introduction

1.1. Introduction to the aims of this thesis

Our decisions are often made within social contexts, and our behaviour and beliefs are subject to change due to the influence of other individuals or groups (e.g., Apps & Ramnani, 2017; Behrens et al., 2008; Campbell-Meiklejohn et al., 2010; Chung et al., 2015; Cialdini & Goldstein, 2004; De Martino et al., 2017; Garvert et al., 2015; Harrigan et al., 2012; Klucharev et al., 2009; Meyer, et al., 2019; Moutoussis et al., 2016; Nicolle et al., 2012; Park et al., 2017; Perreault et al., 2012; Reiter et al., 2019; Shamay-Tsoory et al., 2019; Suzuki et al., 2016; Thomas et al., 2021; Tunçgenç et al., 2021; Zaki et al., 2011). This change can occur explicitly through conformity, or implicitly through contagion, whereby learning about others makes you more like them (Wheeler, 1966). Whilst multiple studies have explored social influence effects, less is known about why we integrate the beliefs and choices of others into our own beliefs and choices, or about why there are individual differences in susceptibility to this influence.

The overarching aim of this PhD thesis is to investigate individual differences in contagion of value preferences in discounting tasks, using a Bayesian neurocomputational approach (largely influenced by the neuroeconomics literature). Within this thesis, individual differences in social influence (i.e., contagion) are explored between groups of neurotypical (NT) and autistic participants (**chapter 3**), on a continuum (as assessed by a measure of autistic traits; **chapters 3** and **4**), between tasks (**chapter 4**) and in both lab-based (**chapters 3**, **4**, and **5**) and online (**chapter 6**) samples.

1.2. Social cognition and social influence

The broad definition of social cognition refers to the cognitive processes which are involved in interpreting the self, others, and social interactions (Forbes & Grafman,

2010; Macrae & Bodenhausen, 2000; Van Overwalle, 2009). Social cognition in humans involves cognitive processes that help us to understand the beliefs and intentions of others, which facilitates successful social interaction and communication (Baron-Cohen et al., 1985). The ability to understand others, and to understand that one's own mental states may differ from those of others, is referred to as Theory of Mind (ToM; Baron-Cohen et al., 1985; Premack & Woodruff, 1978). A discussion of the usage of ToM tasks in autism research is presented below in **section 1.7.**

Our beliefs, behaviours and preferences can be impacted by the beliefs, behaviours and preferences of others, through social influence. This refers to the change in an individual's beliefs or behaviour that is produced by the intentional or unintentional influence of another individual or group (Cialdini & Goldstein, 2004). Two key types of social influence are conformity and contagion. Under conditions of conformity, group pressures result in a change in the behaviour of an individual (Asch, 1956). Under conditions of contagion, an individual learns about the behaviour of another individual or group, and spontaneously imitates this behaviour, or changes their own behaviour to become more similar to that of the other (Ogunlade, 1979; Wheeler, 1966). Social influence effects can be observed in a number of decision-making and valuation tasks (Apps & Ramnani, 2017; Behrens et al., 2008; Bennett, 2015; Campbell-Meiklejohn et al., 2010; Chung et al., 2015; Cialdini & Goldstein, 2004; De Martino et al., 2017; Garvert et al., 2015; Harrigan et al., 2012; Klucharev et al., 2009; Meyer et al., 2019; Moutoussis et al., 2016; Nicolle et al., 2012; Park et al., 2017; Perreault et al., 2012; Reiter et al., 2019; Shamay-Tsoory et al., 2019; Suzuki et al., 2012; Suzuki et al., 2016; Zaki et al., 2011). Research also indicates that our mental health can be affected by social influence (Eisenberg et al., 2013), and recent studies indicate that an individual's

propensity to follow social distancing measures and wear masks during the COVID-19 pandemic is affected by the behaviour of others (Tunçgenç et al., 2021). Thus, social influence is an important area of research, with far reaching consequences.

The social brain is said to underpin social cognitive processes, including social influence (Amodio & Frith, 2006; Beer & Ochsner, 2006; Lee & Harris, 2013; Van Overwalle, 2009). Adolphs (2008) defined the social brain as including the mirror neuron system, (ventrolateral prefrontal cortex; vlPFC, and parietal lobe), mentalizing network (temporo-parietal junction; TPJ, posterior superior temporal sulcus; pSTS, temporal poles, posterior cingulate cortex, and dorsomedial PFC; dmPFC), empathy network (anterior cingulate cortex; ACC, and insula), and reward network (ventromedial PFC; vmPFC, orbitofrontal cortex; OFC, and amygdala). However, there is overlap in the processing of both social and non-social information in many of these areas (e.g., Boorman et al., 2009; Fan et al., 2011; Kable and Glimcher, 2007; Lamm et al., 2011; Ruff & Fehr, 2014), suggesting a lack of specificity for processing social information alone (Lockwood et al., 2020). Although evidence does point to specificity in areas including the gyral portion of the ACC (ACCg) in coding social versus non-social information (Apps et al., 2013; Apps et al., 2016; Apps & Ramnani, 2014; Balsters et al., 2017; Behrens et al., 2008; Hill et al., 2016; Joiner et al., 2017; Lockwood, 2016; Lockwood et al., 2018). Further details of the function of the ACCg is outlined in the next section.

1.3. Decision-making, reward, and learning

Decision making theory makes three key assumptions: (1) there are options to choose between, (2) that choices are made in a non-random way, and (3) that decisions

are goal-directed (e.g., to gain reward or avoid punishment). Our decisions are made based on value judgements, and we tend to choose the option with the highest subjective value, which is continuously updated. From a neurobiological perspective, an object has a subjective value if it acts as a reward or a punisher (Glimcher & Fehr, 2014), and this subjective value can be altered through social influence (Campbell-Meiklejohn et al., 2010).

A decision-making and reward framework which can aid in the indexing of social influence effects is reinforcement learning (RL). This is an associative learning model in which a behaviour is reinforced through reward, or punishment. The discrepancy between what the learner predicts will happen, and what actually happens drives learning, until the learner's prediction is accurate. The discrepancy between predicted and actual outcomes is signalled as a prediction error (PE), which drives learning by modifying representations, expectations and behaviour. Larger PEs drive faster learning and value adaptation, and PE decreases in magnitude as an individual learns, and reward is maximised or punishment is avoided (Doya, 2007; Lee & Harris, 2013; Schultz et al., 1997; Schultz & Dickinson, 2000; Sutton & Barto, 1998; Rescorla & Wagner, 1972). For example, if you were to move to a new area, and there were several local coffee shops, you might visit the highest rated one first, and expect to receive a nice coffee. However, if it wasn't what you expected, and the coffee was too bitter, or you get given almond milk instead of oat milk, you would get a large PE, and lower your own valuation of that coffee shop. If you pop back another time on your way past when all the other shops are closed, and this time you actually get a great coffee, you'd get another PE, and your own subjective valuation for that shop would get a little higher. Hopefully, after popping in

enough times, you realise that the first coffee wasn't representative, and you find your new favourite coffee shop.

Evidence suggests that PE signals are encoded in areas including the ventral tegmental area (VTA; Schultz et al., 1997), ventral striatum, and areas of the mPFC including the OFC and ACC (e.g., Amiez et al., 2005; Apps et al., 2013; Apps & Ramnani, 2014; Balsters et al., 2017; Baram et al., 2021; Behrens et al., 2007; Behrens et al., 2008; Croxson et al., 2009; Ito et al., 2003; Kable & Glimcher, 2007; Lockwood et al., 2018; Reynolds & Wickens, 2002).

Several of these areas are also involved in processing task-related information in social contexts, and thus may be key areas for signalling value update in tasks assessing contagion. Signals in the ventral striatum have been found to code expected (i.e., anticipated) rewards (Baram et al., 2021; Reynolds & Wickens, 2002; Schultz, 2006), and these striatal reward signals have been found to drive learning in line with RL theory (Garvert et al., 2015; Reynolds & Wickens, 2002). However, other research also suggests that this area is not only involved in PE signalling, and is involved in coding signals that are relevant to the task at hand, such as behavioural update (Li & Daw, 2011), evaluating effort exerted for rewards (Botvinick et al., 2009) and timing of rewards independent of value (Klein-Flügge et al., 2011). In social contexts, activity in the ventral striatum has also been found to underpin changes in judgements about facial attractiveness following social information (Klucharev et al., 2009), ratings of music pieces after exposure to the opinions of expert reviewers (Campbell-Meiklejohn et al., 2010), and to signal alignment with group choices (Tomlin et al., 2013). Reward processing in this area has also been found to be affected by social distance, with greater activation in response to feedback for in-group members when participants were asked questions about both in-group and

out-group members (Powers et al., 2016), and greater activation to the receipt of rewards that were shared with friends versus those that were shared with confederates or computers (Fareri et al., 2012). Together, these findings indicate that the ventral striatum plays a crucial role in reward processing and value update in both social and non-social contexts.

The ACC has been implicated in valuation of rewards (Behrens et al., 2007; 2008), and conflict monitoring (Botvinick et al., 2004). Behrens and colleagues (2007) found that variations in ACC signal predicted learning rate, with larger initial signals predicting faster learning, in line with RL. Whilst there is overlap in the areas involved in social and non-social learning, research has also found that separate areas of the ACC encode PEs for social and non-social information, with weighting assigned to both types of information subject to learning and continual update (Behrens et al., 2008). PEs for self-related deviations from expected outcomes have been found in the sulcus of the ACC (ACCs), and social PEs, which reflect discrepancies between the actual and expected outcomes of others have been found in the ACCg (Apps et al., 2013; Apps et al., 2016; Apps & Ramnani, 2014; Balsters et al., 2017; Behrens et al., 2008; Hill et al., 2016; Joiner et al., 2017; Lockwood, 2016; Lockwood et al., 2018). This area also drives prosocial learning, with stronger activation in those who are higher in empathy (Lockwood et al., 2015; Lockwood et al., 2016). Research has also found differences in social PEs in the ACCg in autistic participants versus NT participants, and the extent of these differences was correlated with measures of social deficits in autism (Balsters et al., 2017), providing further support for the role of the ACCg in the processing of social rewards in decision-making tasks.

Whilst a variety of tasks can be used to explore decision-making, reward processing and social and non-social learning, throughout this thesis I focus on exploring contagion of value preferences in neuroeconomics tasks.

1.4. Neuroeconomics and discounting tasks

Core neuroeconomics focuses on the processes connecting sensation and action by examining the neurobiological mechanisms behind sensory decisions. Extended neuroeconomics (which applies to the tasks used in this thesis) combines aspects of economic theory, psychology, and neuroscience and examines the brain mechanisms underlying our choices to form a unified theory of decision-making, and attempt to understand the real reasons behind our behaviour (Glimcher & Fehr, 2014). Neuroeconomic theories can be tested using neuroeconomics games or tasks such as the dictator game, the ultimatum game (Zheng & Zhu, 2013), and discounting tasks, such as temporal discounting (TD; Garvert et al., 2015), and probability discounting (PD; Suzuki et al., 2016). Throughout this thesis, I focus on exploring choice behaviour in TD (**chapters 3**, **4**, **5**, and **6**) and PD (**chapter 4**) tasks.

TD, which is also referred to as delay discounting or intertemporal choice, refers to the decrease in subjective reward value as a function of increasing time (Basile & Toplak, 2015). For example, if an individual is asked whether they would prefer to receive £10 now, or £10 next week, most will prefer to receive the money immediately. Both amounts are the same but waiting until next week to get £10 devalues the reward, so the immediate option (i.e., £10 now) has higher subjective value. The question becomes more interesting when we ask whether someone would prefer to receive £10 now, or £15 next week. Some individuals would be happy to receive the money

immediately, whereas others would prefer to wait a week to get the £15. In TD tasks, participants complete multiple trials with similar choices (i.e., between smaller immediate rewards, and larger later rewards), and differences in choices between participants are expressed as discounting parameters (Basile & Toplak, 2015). TD data can be explained by several different models, including exponential and hyperbolic (Alexander & Brown, 2010; McKerchar et al., 2009):

(1) Exponential: $V = Ae^{-kD}$

(2) Hyperbolic: $V = \frac{A}{1+kD}$

In equations (1) and (2), *A, D,* and *V* represent the amount, the delay attached to the reward, and the subjective value respectively. In equation (1) Euler's number is represented by *e*. An individual fitted parameter referred to as the subject's *k*-value can be derived, and is expressed in both equations as *k*. The exponential function expressed in equation (1) is time consistent and suggests that rewards are devalued at a steady rate over time. The hyperbolic model expressed in equation (2) is time inconsistent and assumes that discounting rates decline over time (Ainslie, 1975; Basile & Toplak, 2015; Green et al., 1997). The commonly observed response pattern in TD tasks indicates that the rate at which rewards are subjectively devalued as a function of delay slows down as delay increases (Ainslie, 1975), which is well captured by a hyperbolic model (Blackburn & El-Deredy, 2013; Garvert et al., 2015; Green et al., 1997; Ohmura et al., 2006).

Throughout this thesis, a log hyperbolic model is used to both calculate the choices of participants, and model the preferences of other agents (as per Garvert et al., 2015). A subject's *k*-value can be thought of as their individual sensitivity to delay. The log *k* parameter values on the TD task were set between -4 and 0 throughout this thesis. A log *k* value of -4 indicates that the individual is barely sensitive to delay and bases their

decision solely on reward value. As log *k* approaches 0, individuals are more sensitive to delay tend to select small immediate rewards, discounting larger later rewards (Carlisi et al., 2017). As well as *k* value, indifference points have been studied as dependent measures in TD tasks. The point at which both choices options (i.e., the small immediate reward and the larger delayed reward) are equally preferred is referred to as the indifference point (Basile & Toplak, 2015).

PD tasks assess preference for risk and involve a series of choices between small, guaranteed rewards, and larger risky rewards, or gambles (e.g., *"Would you prefer to accept £10 or take a gamble for a 50% chance of getting £20?"*). Typically, risk seeking behaviour declines linearly over time from childhood to adulthood, and the majority of adults tend to be risk-averse (Paulsen et al., 2012). Much like TD, PD preferences can be explained by hyperbolic (Green & Myerson, 2004; Myerson et al., 2003; Ohmura et al., 2006) and exponential models (Suzuki et al., 2016), although there is more variability in which model is the most commonly used, and exponential and linear models have also been used to explain PD data (Suzuki et al., 2016). The discounting parameter used to express PD preferences within this thesis (**chapter 4**) is α.

At present, it is unclear whether discounting preferences are domain specific, or domain general, and share a common framework across tasks. Previous research has found correlations between *k* and α values in rhesus macaques, and those that were more patient (i.e., selected more delayed rewards) in a TD task were also more risk-seeking in a PD task (Hayden & Platt, 2007). However, no such correlation in discounting parameters across task was found in human participants (Ohmura et al., 2006). Neuroimaging findings also suggest both task-related differences and similarities, with

the OFC and ventral striatum tracking stimulus value independent of task (Bartra et al., 2013; Levy & Glimcher, 2012; Peters & Büchel, 2009; consistent with evidence that the ventral striatum codes information relevant to the task at hand), and regions in the dmPFC selectively activated by TD, and regions in the parietal lobe selectively activated by PD (Peters & Büchel, 2009). **Chapter 4** is focused on exploring the differences and similarities between the two discounting tasks (and contagion on these tasks) in a behavioural study.

1.5. Social influence in neuroeconomics tasks

Despite previous research indicating that TD and PD preferences are relatively stable across a three-month period (Ohmura et al., 2006), social influence effects on TD and PD preferences have been found, with this change in behaviour underpinned by activation in brain areas implicated in reward processing, decision-making, and social cognition (Apps & Ramnani, 2017; Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Reiter et al., 2019; Suzuki et al., 2016).

In a previous study exploring value representation of TD preferences for the self and for another agent, subjects with low and high rates of TD were paired, and these subjects made choices on behalf of each other (Nicolle et al., 2012). In this study, the researchers found a functional dissociation between different areas of the mPFC. When making choices for themselves, vmPFC activity reflected subjects own value signals, and dmPFC reflected the subject's estimate of the other's preferences. However, the functional roles of these areas switched when the individual made choices for the other. In this instance, activity in vmPFC reflected the value signals of the other, and dmPFC activity reflected the subjects' own value preferences. Behaviourally, participants

choices also became more similar to those of their testing partner. However, this finding of a dissociation between areas of the mPFC has not been found in later research exploring social contagion in TD, although an effect of social contagion has been established. Garvert and colleagues (2015) used computational behavioural modelling to model the behaviour of the other participants with log *k* values ±1 that of the participant, and suggested that differences between self and other value preferences (reflected in a PE signal in ventral striatum) modulate the individual's own internal value representation (i.e., plasticity within the vmPFC). The extent to which plastic changes occurred within the vmPFC corresponded to the strength of the contagion effect.

Piva and colleagues (2019) showed that learning the behaviour of others in a TD task was best explained by a computational model using separate *k* values for self and other, whereas choice behaviour on a PD task was best explained by a model with a shared α value for self and other, indicating that preferences of others are learnt in a task-specific manner. Whilst comparisons between contagion effects in TD and PD tasks have yet to be made, an effect of contagion on PD preferences has also been found, with PE-like signalling found in the dlPFC, modulated by the striatum (Suzuki et al., 2016). Recent behavioural research has also found contagion of PD preferences in adolescent and adult samples (Reiter at al., 2019), with stronger contagion effects found in the adolescent sample for same-age peers than non-age peers. Whilst an effect of social distance was apparent in this sample, it is currently unclear whether contagion of discounting preferences is a social effect, or whether behavioural change occurs as a result of a general informational effect of learning about another agent (regardless of agent identity). In the study by Garvert and colleagues (2015) participant's behaviour shifted to a similar degree when they were told they were making choices on behalf of a

computer agent, as when they were told they were making choices on behalf of a human partner. In this thesis, I examine whether contagion is affected by whether or not participants believe they are making choices on behalf of a human partner across multiple studies (**chapters 3**, **4**, **5**, and **6**), and also whether contagion is affected by social distance (**chapter 5**).

Whilst Garvert and colleagues (2015) found that participants' own discount rates shifted towards the rates of the other agent, this was averaged across directions (i.e., ±1 log *k*). More recent research has found a bias towards becoming more patient (i.e., increasingly negative log *k*; Moutoussis et al., 2016). Slightly larger contagion effects have also been found for risk-seeking (i.e., increasingly negative log α) others versus risk-aversive others (Suzuki et al., 2016), although this difference did not reach significance. However, in this study, the behaviour of the risk-seeking and risk-averse agents was fixed, and more participants in this sample were more risk-averse than risk-seeking, in line with the findings of Paulsen and colleagues (2012). As such, this means that there was a larger difference between participants' average behaviour and the risk-seeking agent than the risk-aversive agent, allowing for a larger shift in behaviour (Suzuki et al., 2016). This highlights the importance of varying behaviour in line with the behaviour of participants, which controls for individual differences in participant's own discounting preferences by keeping constant the distance between participants' own preferences, and the preferences of the other agent(s).

In this thesis (and in Garvert et al., 2015), the behaviour of the other agents is varied in line with the behaviour of participants and is calculated within the experimental scripts. Analyses are also conducted separately for patient and impulsive others (**chapters 3**, and **4**) and for risk-seeking and risk-aversive others (**chapter 4**). The effect

of contagion on PD (**chapter 4**) and TD (**chapters 3**, **4**, **5**, and **6**) preferences is explored behaviourally within this thesis, and the effect of contagion on TD preferences is also explored in an fMRI experiment (**chapter 5**).

1.6. Bayesian methods of exploring social influence

Previous research indicates that Bayesian associative processes are involved in both learning about features of the environment (Behrens et al., 2008; Bennett, 2015), as well as learning of more complex social values (Behrens et al., 2008; Campbell-Meiklejohn et al., 2010; De Martino et al., 2017; Perreault et al., 2012; Suzuki et al., 2012). Social influence can also be explained using Bayesian belief updating, which describes how an individual's prior belief is integrated with new information in order to form a new posterior belief (Bennett, 2015; O'Reilly et al., 2012; Vilares & Kording, 2011). An individual's prior belief (i.e., what they initially thought), and beliefs about the new information can be plotted as two Gaussian distributions, which are weighted according to credibility. The credibility of each of these is represented as the precision (or width) of the Gaussian, and the combination of the two beliefs is weighted according to the credibility of each source of information (Bennett, 2015; De Martino et al., 2017; Park et al., 2017; Vilares & Kording, 2011). For example, if an individual's initial confidence in their own judgement is high, they will be less influenced by social information, and if the credibility of the social information is high (e.g., if the social group is large) and their confidence in their own judgement is low, they will be more influenced (Heyes, 2015). Therefore, accurate learning of the behaviours and preferences of others (which would produce more precise estimates of social information) should

produce a stronger contagion effect (Bennett, 2015; O'Reilly et al., 2012; Park et al., 2017; Vilares & Kording, 2011).

An example of Bayesian processes underpinning social influence comes from a study by De Martino and colleagues (2017), in which participants' own ratings of Amazon products were influenced by reviews of those products. Participants in this study made initial judgements about Amazon products (i.e., their prior beliefs), and also rated their confidence in these judgements (i.e., the credibility of their own belief). They were then exposed to social information before repeating the same judgements. Social information was presented in the form of Amazon reviews, with the credibility varied by manipulating the number of reviews, as well as the number of stars. The researchers found that participants updated their ratings in a Bayesian manner, by weighting the two sources of information according to their reliability (i.e., both initial ratings of confidence, and the credibility of the social information). Research also indicates that new social information is integrated into individual's own decisions based on their degree of certainty in their beliefs on a TD task (Moutoussis et al., 2016), and that the credibility of social information determined by group size also impacts upon jury decisions for criminals (Park et al., 2017).

Bayesian models of social influence have been found to outperform other computational models of social influence because these additionally incorporate information about the credibility of external sources of information (De Martino et al., 2017; Park et al., 2017), as well as the participant's degree of uncertainty in their initial beliefs (Moutoussis et al., 2016). The change from prior to posterior belief is well captured by Kullback-Leibler divergence (D_{KL}), which and accounts for both the amount

of change, as well as the credibility of both sources of information (Kullback & Leibler, 1951).

Forming and updating beliefs depends on the creation of precise prior beliefs about the world, and attaching precise likelihoods to sensory information, which reflects the credibility of this information according to the participant (Bennett, 2015; O'Reilly et al., 2012; Park et al., 2017; Vilares & Kording, 2011). Multiple studies have proposed that differences in Bayesian inference and creation of these priors and likelihoods underlie differences observed in schizophrenia (e.g., Corlett et al., 2009; Fletcher & Frith, 2008; Friston, 2005; Stephan et al., 2006) and autism (e.g., Brock, 2012; Karvelis et al, 2018; Lawson et al., 2014; Lawson et al., 2017; Palmer et al., 2017; Pellicano & Burr, 2012; Van de Cruys et al., 2014) relative to NT individuals. Thus, it is of interest to determine whether differences in Bayesian inference may contribute to differences in contagion in these groups. However, one of the most recent studies found only attenuated sensory likelihoods in autism, and no such effect related to schizophrenia (Karvelis et al., 2018). In this thesis, the focus is on the update of beliefs in autism (**chapter 3**), and in association with autistic traits (**chapter 3** and **4**).

1.7.Reward processing, social cognition, and contagion in autism

Autism Spectrum Conditions (ASC) are characterised by differences in social behaviour and communication (relative to NT individuals), and restricted and repetitive behaviours and interests (American Psychiatric Association, 2013). Throughout this thesis, I refer to autistic individuals, or individuals on the autism spectrum, given that recent research indicates that the majority of autistic individuals prefer these terms, versus person-first language (i.e., person with autism; Botha et al., 2021; Bury et al.,

2020; Kenny et al., 2016). In this thesis, I explore the differences and similarities in contagion of TD value preferences between NT and autistic samples (**chapter 3**), and also whether contagion may be indexed by differences in autistic traits in NT samples (**chapters 3** and **4**).

Differences in the processing of both social and non-social rewards have been observed in autistic participants relative to NT controls (e.g., Delmonte et al., 2012; Dichter et al., 2011; Gonzalez-Gadea et al., 2016; Kohls et al., 2013; Lin et al., 2012), although research into discounting preferences in autism is mixed. Some studies suggest that autistic participants discount future rewards more steeply than NT controls on a TD task (Carlisi et al., 2017; Chantiluke et al., 2014), whereas others find no group differences in discounting (Antrop et al., 2006; Demurie et al., 2012). In a study conducted by Warnell and Redcay (2019), NT and autistic participants were asked to make TD choices on behalf of themselves, and on behalf of four different individuals listed by the participant: (1) the person with whom they were closest, (2) a person they were still close to, but less close than with the first person, (3), a person they knew *"kind of well"*, and (4) a person whom they had met, but did not know well. The researchers found that autistic participants discounted future reward for both themselves and others that they felt close to (i.e., persons (1) and (2)) more steeply than did NT participants. Carlisi and colleagues (2017) also found that activation in the ACC and vmPFC was reduced in autistic participants relative to NT controls during a TD task, consistent with the findings of Balsters and colleagues (2017), who found an absence of social PEs in the ACCg of autistic participants. As adequate learning of the preferences of others is assumed to be required for contagion of preferences to occur (Garvert et al., 2015; Wheeler, 1966), and PE signalling underpins learning (Doya, 2007; Lee & Harris, 2013;

Schultz et al., 1997; Schultz & Dickinson, 2000; Sutton & Barto, 1998; Rescorla & Wagner, 1972) it is possible that contagion of discounting preferences may be weaker in autistic individuals than in NT individuals. At present, previous research into social influence in autism has focused on conformity instead of contagion, and this is mixed. Some studies have found lower rates of conformity in autistic versus NT participants (Van Hoorn et al., 2017; Yafai et al., 2014), whereas others found comparable levels of conformity between groups (Bowler & Worley, 1994; Lazzaro et al., 2018; Van Hoorn et al., 2017).

Autistic individuals have also been found to have difficulties with understanding the perspectives of others, which may contribute to differences in learning about the preferences of others. This is often evidenced by differences in performance between autistic and NT individuals on ToM tasks (Baron-Cohen et al., 1985). Whilst ToM has provided an influential account of social functioning in autism, this has also been problematic (Schurz et al., 2014; Senju et al., 2009; Warnell & Redcay, 2019), and an increasing number of studies have shown that the performance of autistic individuals on ToM tasks depends on multiple factors, including: comorbidities, task framing, and motivation (Frith & Happé, 1994; Hamilton, 2009; Keysar et al., 2003; Oakley et al., 2016; Peterson et al., 2013; Senju et al., 2009; Shamay-Tsoory et al., 2019). It is unclear whether these results reflect differences in ToM processes in autistic individuals, or variability in the way ToM is assessed, given that an increasing number of studies have shown that ToM performance may not be reliable even in NT samples (Schaafsma et al., 2015; Schurz et al., 2014; Warnell & Redcay, 2019). The range of tasks used to assess ToM have also been found to activate different brain areas, and thus it is unclear which processes these are accessing (Schurz et al., 2014; Schurz & Perner, 2015). Given the

increasing variability of ToM performance in both autistic and NT populations, it may be more useful to revisit social processing in autism using a clear neurocomputational framework that precisely isolates the distinct mechanisms underpinning task performance.

It is also possible that differences in the generation of priors and precise likelihoods of sensory information found in autistic versus NT participants (Brock, 2012; Karvelis et al, 2018; Lawson et al., 2014; Lawson et al., 2017; Palmer et al., 2017; Pellicano & Burr, 2012; Van de Cruys et al., 2014) may contribute to differences in how autistic and NT individuals are influenced by the choices of others in discounting tasks. Predictive coding accounts of autism have suggested that autistic individuals generate less precise prior beliefs about the world, and rely more heavily on external sources of information, exhibiting differential weighting of PEs attached to new sensory or environmental input (Lawson et al., 2014; Lawson et al., 2017; Pellicano & Burr, 2012; Van de Cruys et al., 2014). A recent study by Lawson et al. (2017) found overlearning of environmental volatility in autistic adults relative to NT controls, resulting in less surprise at unexpected versus expected outcomes. Less precise prior expectations and more precise sensory representations have also been found in NT individuals with higher levels of autistic traits (Karvelis et al., 2018). In social contexts, the implicit nature of social cues makes it more difficult to generate robust internal priors about others' beliefs as these external sources of information are inherently noisy and less precise. Differential weighting of priors and newly presented social information in autistic versus NT participants may therefore contribute to differences in the integration of these two sources of information to form a new posterior belief (Palmer et al., 2015).

1.8. Summary and outline of thesis

Individual differences in contagion of discounting value preferences requires further exploration. Contagion effects have been observed in several economic decision-making tasks, both in neuroimaging (Campbell-Meiklejohn et al., 2010; De Martino et al., 2017; Garvert et al., 2015; Harrigan et al., 2012; Nicolle et al., 2012; Suzuki et al., 2016), and behavioural studies (Harrigan et al., 2012; Reiter et al., 2019; Thomas et al., 2021). Findings from these studies indicate that key reward areas, including the mPFC, ACC, and ventral striatum, are crucial to the process of contagion (Apps & Ramnani, 2017; Garvert et al., 2015; Nicolle et al., 2012; Suzuki et al., 2016).

At present, no research has explored social contagion effects on economic choice preference in autistic individuals, and research into social conformity indicates that there may be differences in the ways that autistic individuals conform to group norms, relative to NT controls (Bowler & Worley, 1994; Yafai et al., 2014). Autistic individuals may struggle to understand the perspectives of others (Balsters et al., 2017), and differences have also been found in the generation of priors, and precise sensory likelihoods in autistic versus NT individuals (Brock, 2012; Karvelis et al, 2018; Lawson et al., 2014; Lawson et al., 2017; Palmer et al., 2017; Pellicano & Burr, 2012; Van de Cruys et al., 2014). Together, these findings may contribute to differences in learning about the value preferences of others, as well as contagion of value preferences, in NT and autistic individuals.

Across the four experimental chapters presented in this thesis, individual differences in contagion of value preferences are explored in TD and PD tasks. The four experimental chapters each explore different aspects of differences in contagion, and

relationships with other tasks. The methods used throughout the thesis are presented and discussed in **chapter 2**.

In **chapter 3**, contagion of TD preferences is explored in two samples of NT adults, and one sample of autistic adults. In **chapter 4**, the focus is on determining whether there are similarities and differences between task parameters and contagion in TD and PD tasks in a NT sample, and whether autistic traits in this sample are associated with contagion. As a large proportion of participants in these studies did not believe that they were making choices on behalf of real participants, and this did not affect the magnitude of contagion (in line with previous research by Garvert et al., 2015), **chapter 5** explored the effect of social distance on contagion. The study presented in this chapter is an fMRI study, which aimed to determine whether contagion of TD preferences, and learning about other agents, varies depending on the identity of the agent (i.e., close versus distant). This study also explored whether differences in learning and contagion were reflected in blood oxygen level dependent (BOLD) activation in any brain areas, and in regions of interest identified by Garvert and colleagues (2015). **Chapter 6** is an online study, which aimed to explore the relationship between contagion of TD preferences, and advice-taking in a probabilistic learning task. Finally, in **chapter 7**, I present a summary and interpretation of the findings of the four experimental chapters of this thesis.

Chapter 2.

General Methods

2.1. Overview

In this chapter, I summarise and evaluate the key methods used to address the main research goal of the thesis, which is to explore individual differences in contagion of value preferences in discounting tasks. As outlined in the introduction, our preferences on discounting tasks can be altered due to the influence of others (Apps & Ramnani, 2017; Garvert et al., 2015; Nicolle et al., 2012; Suzuki et al., 2016), although little research has explored individual differences in this effect.

Across the four experimental chapters, a Bayesian method of indexing contagion is used, and the majority of the behavioural analyses in each chapter are also accompanied by Bayesian statistics in order to provide further support for the findings. In all of the chapters, the preferences of the other agents are modelled, in line with the preferences of the participant. In **chapter 3**, I present three experimental studies in which contagion of value preferences on a temporal discounting (TD) task is explored in two samples of neurotypical (NT) adults (collected at two separate research sites), and one sample of autistic adults. Within this chapter, contagion is compared across groups (i.e., between the sample of autistic participants and a matched sample of NT participants), and on a continuum of traits (as assessed by the Autism Quotient; AQ). In **chapter 4**, the relationship between contagion and autistic traits is also explored in a NT population, in both a TD and a probability discounting (PD) task. Within this chapter, I also focus on assessing the similarities and differences between contagion and discounting in the across the two tasks. In **chapter 5**, I explore the effect of social distance on contagion of TD preferences, using real-world pairs of close and distant others (i.e., friends or partners, and confederates), although the behaviour in this task is still modelled. This chapter presents both neuroimaging (i.e., fMRI) and behavioural findings. Finally, in

chapter 6, I conducted an online study to explore the relationship between contagion in a TD task, and the propensity to follow social advice in a probabilistic reward learning task.

2.2. Bayesian methods

Bayes' rule is used for calculating conditional probabilities (i.e., the probability of information being correct, given conditions). According to Bayes' rule, the likelihood of new information is integrated with priors (e.g., our own prior beliefs in the context of Bayesian belief updating), by weighting this likelihood by the strength of the prior to form a new posterior belief (Bennett, 2015; Knill & Pouget, 2004; Mathys et al., 2011). Essentially, when applied to Bayesian belief updating (i.e., the updating of our own beliefs), Bayes' rule helps us to decide how much to change our minds, given the available evidence (i.e., what we already know, and the information we have already been presented with). If we are more confident about our prior beliefs or experience (i.e., precise prior beliefs), less weighting will be placed on new information, and the update in belief will be small. If we are less confident, more weighting will be placed on this new information, and the belief update will be larger. Indeed, Bayesian associative processes have been found to underlie learning in social and non-social contexts (Behrens et al., 2008; Bennett, 2015; Campbell-Meiklejohn et al., 2010; De Martino et al., 2017; Perreault et al., 2012; Suzuki et al., 2012). In the context of using Bayes rule to calculate evidence for or against a model, Bayes rule is used to quantify the evidence in favour of that model by integrating prior information, with new information, such as data.

2.2.1. Bayesian methods of hypothesis testing

Null hypothesis testing, in which the probability (expressed as a *p*-value) of observing this value or a more extreme result is computed alongside test statistics such as F-values, is popular, although it also has limitations. One of the most important limitations is that using *p*-values alone does not tell the reader how much the hypothesis of interest is favoured versus the null-hypothesis. This is because *p*-values are not used to quantify evidence in favour of the null, and are used only to disprove the null hypothesis. Whereas Bayes factors also account for the strength of the effect according to the data (e.g., the number of observations), *p*-values are also unable to do this alone (Dienes, 2014).

Bayes factors can be used to compliment *p*-values and effect sizes in order to provide the reader with more information about the strength of the effect. Bayes factors can do this by quantifying the available evidence (i.e., given the dataset) in favour of one hypothesis versus another, given a prior probability. For example, BF_{10} is used to quantify the evidence for the hypothesis of interest versus the null, given the data, and the inverse of this (i.e., BF_{01}) is used to quantify the evidence in favour of the null hypothesis versus the hypothesis of interest, given the data (Kass & Raftery, 1995; Quintana & Williams, 2018). In both cases, the higher the Bayes factor, the stronger the evidence for that hypothesis, versus the other.

Within this thesis (**chapters 3**, **4**, **5**, and **6**), Bayes factors are presented alongside *p*-values and effect sizes for the majority of analyses, in order to provide additional support for findings. Both BF_{10} and BF_{01} are used throughout, with this reporting dependent on significance. BF_{10} is used when the *p*-value is below the critical value of $p < .05$, and BF_{01} is used when the *p*-value is > .05. The exception to this is when

Bonferroni corrections are made to *p*-values to allow for multiple comparisons (as in **chapter 4**). Throughout this thesis, jsq package (in Jamovi) default priors are used to calculate Bayes factors, and these are interpreted in line with the package criteria (as shown in **Figure 1**).

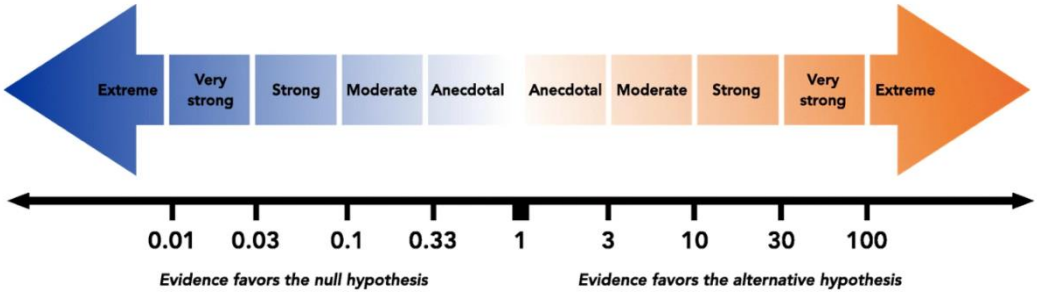


Figure 1. Interpretation of Bayes factors. This shows the interpretation of Bayes factors that is adopted within this thesis (presented for BF_{10}). Figure from Quintana and Williams (2018), and interpretation is from Lee and Wagenmaker (2014, as cited in Quintana & Williams, 2018).

2.2.2. Exploring social influence using Bayesian methods

Bayesian models of social influence incorporate both information about an individual's certainty in their initial beliefs (through the width of priors), and information about the likelihood of newly presented social information (expressed as the width of the Gaussian representing the likelihood). These two sources of information are combined in order to represent the individual's updated belief, as the posterior. Bayesian models of social influence have been found to outperform other computational models of social influence because these models incorporate these two types of information into the calculation of the updated belief (i.e., following social information; De Martino et al., 2017; Moutoussis et al., 2016; Park et al., 2017). The change from prior to posterior belief is well captured by Kullback-Leibler divergence (D_{KL}), which accounts for both the amount of change, as well as the credibility of both sources of information (Kullback & Leibler, 1951). Further examples of the application of Bayesian models to social

influence are presented in **chapter 1, section 1.6.** Prior and posterior beliefs for an example participant from the Study 2 sample presented in **chapter 3** is shown in **Figure 2**.

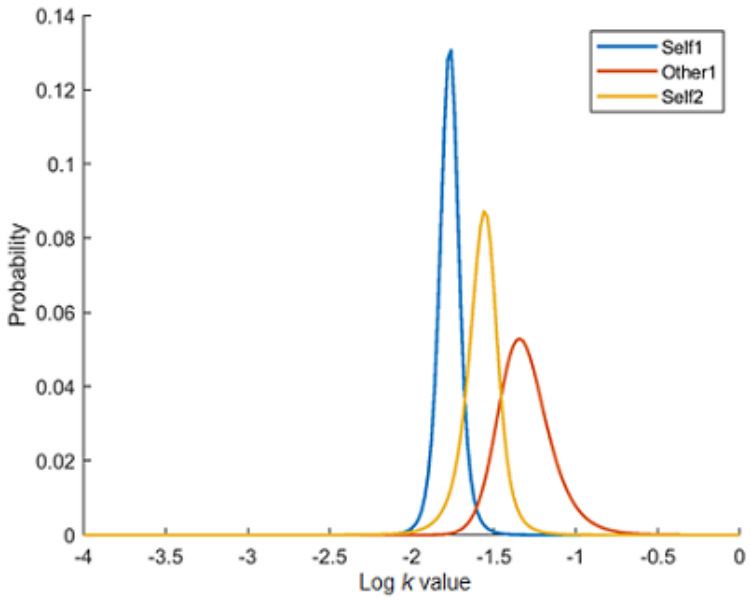


Figure 2. Example participant behaviour. This graph shows posterior probabilities for an example participant's log *k* value at the end of the Self1-, Other1- and Self2-blocks. D_{KL} is used to quantify the divergence between two distributions (i.e., Self1 and Self2, which is after the influence of Other1), and accounts for the width of the Gaussian for both Self1 and Self2. It is apparent from the graph that the posterior for Self2 has shifted away from the posterior for Self1, and towards Other2, which shows contagion.

2.3. Discounting tasks

As outlined in **chapter 1**, the discounting tasks used within this thesis assess decrease in subjective reward value as a function of increasing time (TD), or risk (PD; Green & Myerson, 2004; Myerson et al., 2011). TD tasks involve a series of choices between smaller sooner rewards (e.g., £3 now) and larger later rewards (e.g., £6 tomorrow), and PD tasks involve a series of choices between small, guaranteed rewards (e.g., guaranteed £3) and larger risky rewards (e.g., 50% chance of £6; **chapter 4**). These types of choices are similar to the choices that occur in our everyday lives, for example, it is often important to be patient when working towards goals and delay gratification

(e.g., *"Should I enjoy myself tonight or study for the exam tomorrow?"*; Lempert et al., 2019). Indeed, discounting has also been associated with real world behaviours such as academic performance (Kirby et al., 2005; Wulfert et al., 2002).

These tasks were first explored in human participants by Rachlin and colleagues in 1991, who found that a hyperbolic model, which is time inconsistent and assumes that discounting rates decline over time, best explained both TD and PD, versus exponential models. Although both TD and PD data have since been explained by both hyperbolic and exponential models (Alexander & Brown, 2010; Basile & Toplak, 2015; Blackburn & El-Deredy, 2013; Garvert et al., 2015; Green et al., 1997; Green & Myerson, 2004; McKerchar et al., 2009; Myerson et al., 2003; Ohmura et al., 2006; Suzuki et al., 2016).

2.3.1. Modelling the choices of participants, and other agents

Throughout this thesis, a log hyperbolic model is used for both TD and PD, to both derive the preferences of participants, and model the choices of other agents. As discussed in **chapter 1**, log *k* value is the discounting parameter for TD tasks, and log α is the discounting parameter in PD tasks.

The hyperbolic functions used to model these preferences in TD and PD tasks are as follows:

(1) $$V = \frac{M}{1+kD}$$

(2) $$V = \frac{M}{1+\alpha(1-P/P)}$$

Equation (1) shows the subjective value (*V*) calculated for one of the options on a TD task trial, and equation (2) shows the calculation of subjective value for a choice option on a PD trial. As stated, subjective value in each equation is represented by *V*, *M* is the reward magnitude, *k* or α is the agent's discount rate, *D* indicates the delay period

in days, and *P* is the probability of receiving a reward. Log *k* parameter values were set between -4 and 0 throughout, and log α parameters were set between -2 and 2.

The subjective values of V_{SS} and V_{LL} are transformed into choice probabilities using the following softmax equation:

$$(3) \qquad P_{LL} = \frac{1}{1+e^{-\beta(V_{LL}-V_{SS})}}$$

The subjective value of the immediate (TD) and guaranteed (PD) options is referred to in equation (3) as V_{SS}, and the value of the delayed (TD) and gamble (PD) option is referred to as V_{LL}. The subjective value of V_{SS} always corresponds to the magnitude of the reward (*M*), because the delay period in days is 0 (represented by *D* in equation (1)), or the probability attached to the reward is 0 (represented by *P* in equation (2)). Equation (3) shows the calculation of choice probability for P_{LL} which is the delayed (TD) or gamble (PD) option (the subjective value of P_{SS} can then be calculated as 1 - P_{LL}). In equation (3), β is a free parameter characterising noise in the agent's choices. This value was set between -1 and 1 throughout the experiments in this thesis.

When choices are modelled on behalf of other agents, the log *k*, log α, and log β values are already set (with log *k* and log α set at ± 1 of the participant's own discounting parameter and log β set at 1). The subjective value of each option on each trial can be computed for the other agent by inputting the values into the equations. When deriving the discounting parameters and log β values of participants, subjective value will be estimated for each option according to the option chosen by participants, and parameters can be estimated according to the participant's choice. Parameters (i.e., log *k*, log α, and log β) are then updated on a trial-by-trial basis using a Bayesian update rule. A uniform prior is updated on each trial to form the posterior, which is calculated as the likelihood of an individual's choice given the discounting parameter (log *k* or log α), and log β,

weighted by the prior. As the participant completes more choices, the estimate of their discounting parameters becomes more precise.

2.3.2. Individual differences

Individual differences in discounting tasks can be observed both within and between clinical and non-clinical groups (Amlung et al., 2019; Kekic et al., 2020). For example, increased impulsivity in TD has been associated with substance use disorders (Amlung et al., 2017), schizophrenia (Heerey et al., 2011; Yu et al., 2017), and attention-deficit/hyperactivity disorder (Jackson & MacKillop, 2016). Some studies also find increased impulsivity in autism, as discussed in **chapter 1**. Individual differences in discounting preferences can be quantified using discount rates (Peters & Büchel, 2011). As discount rates (i.e., k for TD and α for PD tasks) are used as dependent measures across multiple studies, this makes findings across studies directly comparable when assessing discounting preferences across studies.

In order to account for individual differences in discounting preferences when examining an effect of contagion, the behaviour of other agents in this task can be modelled so that the distance between the behaviour of the agent and each participant can be held relatively constant. Some of the previous research exploring contagion in discounting tasks has used data from real participants for the choices of the other individual (Nicolle et al., 2012; Piva et al., 2019), or computational modelling to model a fixed (Suzuki et al., 2016), or variable pattern of choices (Garvert et al., 2015). Using choices from real participants or modelling a fixed pattern of choices introduces variability into the data, as the extent to which the choices of the other diverge from those of the subject will vary for each participant, whereas modelling choices in line with the preferences of the participant avoids this. In **chapters 3**, **4**, **5**, and **6**, the discount rates of

the other agents are modelled within the experimental scripts to be ±1 the log *k* or log α of the participant (as per Garvert et al., 2015).

However, whilst this does control for differences in discounting preferences between participants and the other agents, if a participant makes extreme choices on either task (e.g., a participant chooses the immediate option or the guaranteed option on almost every trial), it is hard to model the discount rate of another agent whose behaviour is sufficiently distinct from the behaviour of the participant. This is particularly apparent in the PD task in **chapter 4**, although log *k* or α ±1 is used throughout this thesis in order to keep this distance constant across tasks and experiments. A discussion regarding varying the extent to which the behaviour of the other agent differs from that of the participant is presented in **chapter 4.**

2.3.3. Adapting discounting tasks

Monetary rewards are used throughout the four experimental chapters in this thesis, but discounting tasks can also be easily adaptable to different audiences. For example, the marshmallow test can be used to test delay of gratification (Mischel et al., 1972), and studies in adults have also used preferred snacks as reinforcers instead of monetary rewards (Göllner et al., 2018; Knolle-Veentjer et al., 2008), although non-monetary rewards are often more steeply discounted (i.e., individuals make more impulsive choices) than monetary rewards (Odum et al., 2020). When these tasks are used to assess contagion, different cues can be incorporated into the task to inform participants who they are making choices on behalf of. In **chapters 3**, **4**, and **5** different text colours and different wording is used to denote the identity of each agent, and in **chapter 6**, different wording is presented at the top of each trial to inform participants who they are deciding on behalf of for that trial. Previous research has also used images

of other participants to show who participants are making choices on behalf of (e.g., Reiter et al., 2019). As such, these tasks are easily adaptable to use in a variety of samples. In all of the experimental chapters in this thesis, the trial structure is also held constant, with only the information denoting the identity of the agent varying on each trial. As only these small aspects are changed in order to have social versus non-social trials, these tasks are also appropriate for both behavioural and neuroimaging experiments, as the differences between social and non-social conditions can be easily controlled.

2.4. Probabilistic reward learning task

The probabilistic reward learning task employed in **chapter 6** is a two-forced choice card game with associated reward probabilities that required integration of social and non-social cues (adapted from Behrens et al., 2007; Behrens et al., 2008; Henco et al., 2020; Sevgi et al., 2020). The version of the task used here is presented as a card game on a card table, with a blue card and a red card, each with reward amounts attached. On each trial, one card is the winning card. Importantly, the reward magnitudes were independent of the reward probabilities associated with each card, so that participants could not use the reward magnitudes to infer the probability of each card being the winning card. This task was originally developed by Behrens and colleagues (2008), and was used to determine that the same associative learning processes underpin both social and non-social learning.

Since the task was developed, it has also been used to assess differences in learning of social and non-social information predominantly in NT samples (Henco et al., 2020; Sevgi et al., 2020). One of these studies also compared task performance in

individuals with high and low levels of autistic traits and found group differences in the tracking of social information (Sevgi et al., 2020).

The social component included in the task is **chapter 6** is presented in the form of one of the cards appearing in the bottom right-hand corner of each trial, presented as information chosen by another participant (see **Figure 3** for an example screen from the task used in **chapter 6**). In the original study, the social cue was presented in the form of a box around the recommended choice (Behrens et al., 2008) and research has also used a face with the gaze shifting towards the chosen option (Sevgi et al., 2020). The initial study and more recent research have used explicit advice to inform participants about the social cue (Behrens, 2008; Diaconescu et al., 2014), whereas in the study by Sevgi and colleagues (2020), participants were told that the social information (i.e., the face) was presented to make the visual display more interesting, and they were not told that they needed to learn about whether the advice was helpful or not. In **chapter 6**, participants were informed about the social cue, and were told that the card was chosen by another participant, who was either being collaborative, and presenting helpful information, or competitive, and presenting incorrect information about the rewarded card. This is because participants were also told explicitly in the TD task that they would need to try to predict the choices of the other participant and providing reasoning for the social information across tasks aided in keeping the type of social information constant (i.e., explicit).

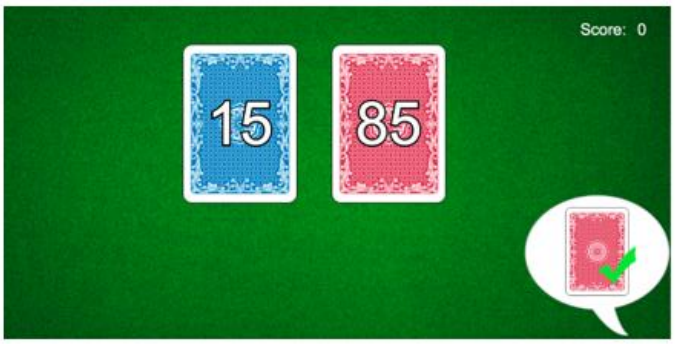


Figure 3. Probabilistic reward learning task experimental screen. This shows an example experimental screen, with reward magnitudes presented on each card, and the social information displayed in the bottom right-hand corner. The participant's score is shown on the top right-hand side, and this increases throughout the task when the participant chooses the correct card, and collects the points from that card.

The probability of both the social information being correct and of the blue card being the rewarded card was manipulated according to probability schedules adapted from Sevgi et al (2020). The probability attached to the two types of information varied independently of each other, so that participants could not infer how likely the social information was to be correct from the most likely to be rewarded card, and vice versa. Both the card reward schedule, and the social reward schedule also had separate stable and volatile phases (i.e., the reward probability associated with the information type was high for a large block of trials in the stable condition, or varied more frequently in the volatile condition). This means that participants can learn about each piece of information (i.e., social and non-social) separately. In a volatile phase, participants who are able to track the information effectively would adjust their decisions quickly, making use of the most recent information to guide their next decisions. In the stable phase, participants who are able to track the information effectively would adjust their decisions slowly, making use of information across multiple trials. Behrens and colleagues (2008) found that participants did not blindly follow the social advice in their task, and they also found no significant effect of participants assuming throughout the task that the social

information would be as informative as it was on the previous trial. This indicates that participants were able to make use of the social information appropriately on this task.

As the task has been used multiple times in previous research, with clear evidence that participants are able to independently learn about social and non-social information (Behrens et al., 2008; Henco et al., 2020; Sevgi et al., 2020), this task is well-placed for examining the relationship between task performance and contagion.

2.5. fMRI

Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging technique that allows the mapping of human brain function. It is able to cover the whole brain, and has high spatial resolution, although comparatively poor temporal resolution. fMRI records a blood oxygen-level dependent signal (i.e., BOLD signal), which relies on the differing magnetic properties of oxygenated and deoxygenated blood. When brain regions are active, these consume and require more oxygen than areas that are inactive, and thus fMRI identifies areas that have more oxygen. However, as the BOLD response relies on movement of oxygen, this is particularly sluggish.

2.5.1. Measuring differences between conditions

Differences in the BOLD signal are generally compared between conditions, with minor aspects of each condition changed, so that changes in BOLD signal can be attributed to individual changes between conditions. In **chapter 5**, BOLD responses are compared between conditions for close versus distant others using paired t-contrasts. In this task, all features of the task are held constant, and the only aspect that changes between the two social conditions is the identity of the other. Designing the task in this manner ensures that differences in the BOLD response between conditions should be

attributed to differences in the social distance of the other agent, providing motion is controlled for. In **chapter 5**, there are also multiple screens presented on each trial (two screens in Self-trials, i.e., Offer and Choice, and three screens in Other-trials, i.e., Offer, Choice, and Feedback), and each of these screens was entered as a condition in the design-matrix (with the timings of the appearance of each screen added to the design-matrix). To ensure that each element of the design-matrix was maximally decorrelated, a variable jitter of 2-5 seconds was presented between trial screens, so that the BOLD response could be time-locked to specific trial events throughout the entire experiment (without contaminating effects of other trial events, e.g., Balsters & Ramnani, 2008; 2011). As the choice options were also presented at both the Offer and Choice screen in both conditions, the position of the options on each of these screens these was independently randomised (i.e., the immediate option presented at the top or bottom of the screen at the Offer stage, and to left or right of the screen at the Choice stage). This ensured that any BOLD activation observed at the Offer stage was unrelated to motor preparation effects from participants preparing their response for the Choice stage.

2.5.2. Noise in the signal

The robust weighted least squares approach (RWLS; developed by Diedrichsen & Shadmehr, 2005) was used to specify and estimate the fMRI design-matrix in **chapter 5**, because this method detects and adjusts for noise artefacts in fMRI data. A large amount of noise in the data is caused by motion effects from participant movements such as eye movements, or physiological effects such as heart beats (Frank et al., 2001) and RWLS is able to identify these artefacts regardless of their origin (Diedrichsen & Shadmehr, 2005). Whilst participants with head movement jumps > 5mm were excluded from all analyses in **chapter 5**, and the realignment process used during data pre-

processing can correct for motion effects, residual noise often remains in the data (Grootoonk et al., 2000). A multiband EPI sequence was also used to collect our data, which reduces the signal to noise ratio, allowing a reduction in the repetition time (TR) and increased BOLD sensitivity (Demetriou et al., 2018; Todd et al., 2016).

Using the RWLS approach allows for further control of noise in the data. RWLS works by estimating the variance of the noise for each image in the time series. This is because small events such as eye blinks, swallowing, or small head movements will only occur in some images, and thus the noise needs to be estimated for each image independently. The noise estimates for each image are then used to obtain a weighted least squares estimate (weighted by the inverse of the image variance) of the regression parameters within a linear model (here, the design-matrix). Diedrichsen and Shadmehr (2005) showed significant variation in noise across images in the fMRI time-series in their test data set, and that the RWLS approach successfully detected noise artefacts in the data (whereas noise in individual images would traditionally only be identified by visually examining the fMRI time series) so that this noise could be accounted for in the general linear model (GLM). Accounting for this noise also improved sensitivity in detecting regions of activation. As the RWLS method is also used in **chapter 5** (alongside excluding participants based on head movements > 5mm and correcting for as much noise as possible during realignment), the data should be maximally sensitive to detect possible differences in mean BOLD signal across conditions using the specified contrasts.

2.6. Assessing autistic traits in the general population

In **chapters 3** and **4**, autistic traits are related to contagion on a continuum in samples of NT participants. Multiple studies have also examined variation in task performance dependent on autistic traits in NT populations (e.g., Goris et al., 2019; Karvelis et al., 2018; Sevgi et al., 2020). In the initial Autism Quotient (AQ) paper (Baron-Cohen et al., 2001), this measure was also used to assess the distribution of autistic traits in the general population, and thus from its development, this measure has been able to provide a measure of traits in those with and without an autism diagnosis. Within this thesis, a Likert scoring method (originally developed by Austin, 2005, in order to retain individual variation in the extent of agreement with each statement) was used to score the AQ. The Likert scoring method is used in **chapters 3** and **4** because these chapters examine the relationships between individual differences in autistic traits, and contagion, and thus a method which retains individual variance in participant scores is preferred. In contrast to the original binary scoring method employed by Baron-Cohen et al., (2001), whereby each item is scored zero or one, regardless of whether the individual selected strong or slight agreement/disagreement with the statement, the Likert scoring method has also demonstrated higher internal consistency and test-retest reliability in a NT sample (Stevenson & Hart, 2017), as tested used here.

Chapter 3.

Contagion of Temporal Discounting Value Preferences in Neurotypical and Autistic Adults

Louisa Thomas
Patricia L. Lockwood
Mona M. Garvert
Joshua H. Balsters

Thomas, L., Lockwood, P. L., Garvert, M. M., & Balsters, J. H. (2021). Contagion of temporal discounting value preferences in neurotypical and autistic adults. *Journal of Autism and Developmental Disorders.* Advance online publication.

Chapter starts on **page 53** of the thesis

51 pages in chapter, including supplemental information and references

Abstract

Neuroeconomics paradigms have demonstrated that learning about another's beliefs can make you more like them (i.e., contagion). Due to social deficits in autism, it is possible that autistic individuals will be immune to contagion. We fit Bayesian computational models to a temporal discounting task, where participants made decisions for themselves before and after learning the distinct preferences of two others. Two independent neurotypical samples (N = 48; N = 98) both showed a significant contagion effect; however the strength of contagion was unrelated to autistic traits. Equivalence tests showed autistic (N = 12) and matched neurotypical samples had similar levels of contagion and accuracy when learning about others. Despite social impairments being at the core of autistic symptomatology, contagion of value preferences appears to be intact.

Introduction

Social influence has been shown to bias our behaviours and preferences (Behrens et al., 2008; Cialdini & Goldstein, 2004; Izuma, 2013; Meyer et al., 2019; Raafat et al., 2009; Shamay-Tsoory et al., 2019). For example, multiple studies have demonstrated that learning the alternative value preferences of another person significantly shifts our own value preferences – a phenomenon referred to as contagion (Apps & Ramnani, 2017; Garvert et al., 2015; Nicolle et al., 2012; Suzuki et al., 2016; Reiter et al., 2019; Wheeler 1966). Given that autistic individuals often struggle to understand the alternative preferences of others (Baron-Cohen et al., 1985; Hamilton, 2009), it is possible that they are immune to the effects of contagion or may experience contagion to a lesser degree than neurotypical (NT) individuals. Here, we use Bayesian modelling to investigate contagion of value preferences in autistic individuals and autistic traits in the NT population.

One computational framework increasingly employed in cognitive and social neuroscience is Bayesian belief updating (Bennett, 2015). Previous research has found that Bayesian associative processes underlie learning about environmental features as well as more complex social values (Behrens et al., 2008; Bennett, 2015; Campbell-Meiklejohn et al., 2010; Mussey et al., 2015; Perreault et al., 2012; Suzuki et al., 2012). Within a Bayesian framework, contagion can be interpreted as an integration of an individual's prior beliefs and their beliefs about newly presented social information (i.e., from another individual or group) to form a new posterior belief. Measuring social influence using a Bayesian framework not only helps us to define our concepts more clearly, but it also allows us to precisely quantify observed behaviours. Statistical model comparisons suggest that Bayesian approaches outperform other computational models

of social influence, by additionally incorporating information about the credibility of external sources of information (De Martino et al., 2017; Park et al., 2017), as well as the participant's degree of uncertainty in their initial beliefs (Moutoussis et al., 2016), using the precision (i.e., the width) of priors.

Bayesian frameworks (also called predictive coding accounts) have also been employed to explain behaviours in autism spectrum conditions (ASC, or autism). Predictive coding accounts of autism suggest that autistic individuals struggle to generate precise internal prior beliefs about the world (both social and non-social; Karvelis et al., 2018; Lawson et al., 2017; Pellicano & Burr, 2012; Van De Cruys et al., 2014). Pellicano et al. (2011) found that autistic children took longer to learn reward probabilities than typically developing children in a foraging task and were less consistent and optimal in their search strategy, indicating a deficit in rule learning. A more recent study by Lawson et al. (2017) found differences in learning of volatility, with autistic adults being more likely than NT adults to overlearn about environmental volatility. This resulted in less surprise when outcomes were unexpected, relative to expected. A decrease in surprise at unexpected outcomes was also related to greater symptom severity. Similar findings have been uncovered in NT individuals with higher levels of autistic traits. Karvelis et al. (2018) explored predictive coding in NT individuals using a visual learning task and showed that increased autistic traits in this sample corresponded to less precise prior expectations, combined with more precise sensory representations. Taken together, these findings suggest a deficit in generating internal priors that form the prediction of reward outcomes in autistic individuals relative to NT controls, and that this finding can also be observed in NT individuals with higher levels of autistic traits. This may translate to decreased learning about the subjective values of others in tasks measuring contagion.

Several studies have shown contagion of subjective value preference within NT samples, using temporal discounting (TD) tasks. In these studies, participants made TD choices on behalf of themselves and others with alternative value preferences. After learning the alternative value preferences of another individual, or group, the participant's own discount rate changed to become more like that of the other (Apps & Ramnani, 2017; Garvert et al., 2015; Nicolle et al., 2012). Research into discounting in autism is mixed. Some studies suggest that TD follows the same pattern as in NT individuals (Antrop et al., 2006; Demurie et al., 2012; Warnell et al., 2019), whereas others have found that autistic individuals discount future rewards more steeply than NT controls (i.e., are more impatient; Carlisi et al., 2017; Chantiluke et al., 2014). To account for individual differences in contagion, the value preferences of the other agents in this study vary in line with the participant's own preferences.

Garvert et al. (2015) suggest that differences between self and other value preferences (reflected in a prediction error (PE) signal in ventral striatum) modulate the individual's own internal value representation (i.e., plasticity within the vmPFC) in the TD contagion task. The extent to which plastic changes occurred within the vmPFC corresponded to the strength of the contagion effect. This is particularly relevant in light of the results of Balsters et al. (2017), who found that autistic individuals do not produce PE signals about others (i.e., a brain signal describing the difference between expected and unexpected outcomes for another person). The absence of this social PE signal could make autistic individuals immune to contagion effects, as appropriate social PE signals should be necessary for learning about another person. In addition, autistic adults have been shown to be more consistent and inflexible in their choices than NT controls in a probability discounting (PD) task (Wu et al., 2018). A preference for predictability and

sameness has also been found in NT individuals with higher levels of autistic traits (Goris et al., 2019). In this study, significantly positive correlations were found between the level of autistic traits and preference for both more predictable music pieces, and for visual items that were increasingly similar to a visual prime. Participants with higher levels of autistic traits were also faster to choose decks of cards with predictable outcomes, indicating an implicit preference for the more predictable decks. Autistic participants have also been found to revert to previously preferred responses more quickly than neurotypical controls in a probability reversal learning task with feedback provided on an 80:20 schedule (D'Cruz et al., 2013). Whilst there were no differences in learning of reward contingencies in this study, autistic participants were also faster than NT controls to revert to the previously rewarded response following incorrect feedback. Together, these studies indicate that the internal value preferences of autistic individuals, and NT individuals with higher levels of autistic traits, may not change due to the influence of the other, or that these participants may revert to their own preferences more quickly after making choices on behalf of another individual.

We conducted three studies to examine: (1) whether NT individuals shift their preferences when they learn about other people's preferences, (2) whether levels of autistic traits in the NT population are associated with contagion of value preferences, and (3) whether autistic adults show comparable shifts in preference to NT controls. In Study 1, participants completed a TD contagion task as well as a measure of autistic traits. In Study 2, we tested a larger replication sample of NT participants at a separate research site. Finally, in Study 3, we tested a sample of autistic adults on the TD contagion task. In all three studies, we conducted separate analyses on contagion for more impulsive and more patient agents, as previous research has suggested contagion

effects are stronger when presented with a more patient agent (Moutoussis et al., 2016) using a similar paradigm.

We predicted that we would observe contagion effects in both the NT samples, such that people would shift their preferences after learning about another person's preferences. We also predicted that levels of autistic traits in the NT samples would be associated with lower levels of contagion, and that autistic adults would be less accurate at learning the value preferences of the other agents. This reduced learning would result in less contagion of value preferences relative to NT adults.

Methods

Participants

Study 1

A sample of 49 NT participants (23 male, 24 female, 2 unreported, mean age = 23.73, SD $\pm$ 3.86) were recruited from the University of Oxford. Participants were paid at a rate of £10 per hour and were told that they would receive an additional bonus based on a randomly selected trial from the experiment. In fact, participants were paid a randomly selected bonus ranging between £1 and £10 on the day and were informed that a trial had been chosen. Informed consent was obtained, and ethical approval was granted by the Departmental Ethics Committee at Oxford.

Study 2

One hundred NT participants (60 female, mean age = 21.35, SD $\pm$ 2.11) volunteered to take part, and were recruited from the Royal Holloway, University of London campus. As an incentive for participation, participants were invited to enter into a prize draw to win an Amazon voucher for the amount of one of their chosen monetary

outcomes (£1-20), which was drawn when data collection commenced. Immediate amounts (e.g., £3 now) were sent immediately, and delayed amounts (e.g., £8 in two weeks) were sent following the specified delay, with a notification of the win sent immediately following the draw. Informed consent was obtained, and Ethical approval was granted by the Royal Holloway Departmental Ethics Committee.

Study 3

A further sample of 14 participants with a diagnosis of an ASC (8 female, mean age = 22.29, SD $\pm$ 3.36) were recruited from the Royal Holloway, University of London campus, and externally. Although this sample is smaller than the Study 1 and Study 2 samples, a power analysis indicated that a sample of four was required to achieve 80% power when examining an effect of contagion in previous discounting tasks. Therefore, the sample of 14 included here maximises power and allows for exclusions and potential dropouts. Participants in this sample were also invited to enter into the same prize draw as the participants in Study 2 to win an Amazon voucher for the amount of one of their chosen monetary outcomes (£1-20). An additional three participants were recruited to this study following the prize draw and were paid £8 for their time. Informed consent was obtained, and Ethical approval was granted by the Royal Holloway Departmental Ethics Committee.

Scores on the autism quotient (AQ; Baron-Cohen et al., 2001), scored using the Likert scoring method (Austin, 2005) are presented in **Table 1** for the final samples for all three study groups, and the Study 1 and Study 2 samples combined (following the exclusions outlined at the end of the **Methods** section). See **Supplemental Table 1** for AQ data scored using the binary scoring method (Baron-Cohen et al., 2001) for all three samples.

Table 1. Descriptive statistics (mean (SD±)) for AQ subscale scores and total scores for both NT samples (Study 1 and Study 2), and the ASC sample (Study 3), following exclusions.

	Study			
	Study 1	Study 2	Study 1/Study 2 combined	Study 3
AQ				
Social Skills	25.87 (3.21)	20.56 (3.99)	22.28 (4.50)	29.83 (4.13)
Attention Switching	26.94 (3.30)	24.46 (3.97)	25.26 (3.93)	33.17 (3.66)
Attention to Detail	25.61 (3.90)	24.85 (4.36)	25.10 (4.21)	29.42 (4.58)
Communication	24.89 (4.54)	19.82 (3.79)	21.47 (4.68)	31.67 (3.89)
Imagination	25.78 (3.82)	19.81 (4.04)	21.75 (4.85)	25.42 (3.83)
TOTAL	129.09 (10.41)	109.51 (13.13)	115.85 (15.34)	149.50 (14.39)

Discounting tasks

Study 1

Participants first completed a TD contagion task. This was presented in MATLAB and was developed using the experimental script programmed by Mona Garvert (Garvert et al., 2015), using the Cogent 2000 v125 graphics toolbox. Participants made a series of self-paced hypothetical choices between two monetary values (£1-20) presented simultaneously, choosing between a small amount of money available immediately (e.g., £3 now), or a larger amount available after a specified delay period (tomorrow, 1 week, 2 weeks, 4 weeks, 6 weeks, 2 months, or 3 months). Responses were made using left and right arrow keys, corresponding to the location of the choice on the screen. The side (left or right) of the immediate option was randomised on a trial-by-trial basis.

The task was divided into five blocks of 50 trials (Self1, Other1, Self2, Other2, and Self3), with a self-timed break after 25 trials. During Self-trials, participants were

instructed to choose for themselves according to their own preferences, as they believed that one of these choices would be selected as their potential bonus payment. In blocks two and four (i.e., Other1-block and Other2-block), participants made choices on behalf of two simulated agents, and were informed that these were the choices of two previous participants. In fact, the behaviour of these two agents was modelled online based on the participants own choices in the first block (i.e., Self1), to be plus or minus one of the participants own discount rate. The direction (±1) of Other1 and Other2 was counterbalanced across participants within the script (see the **Estimation of discount rates and simulation of the Other's choices** section below for details of this calculation). Two gender-matched names (or two randomly selected names for participants who did not report their gender) were chosen to represent these two agents. During all Other-trials, feedback was also displayed on a trial-by-trial basis, to inform participants whether their choice for the other agent was correct. If participants chose the option that would be preferred by this modelled agent, feedback was displayed stating that the choice was correct, or incorrect if the participant chose the other option. This feedback remained on the screen for 1000 ms. Task order and example experimental screens are shown in **Figure 1**. See **Supplemental Information** for details of how choice pairs were generated for all blocks.

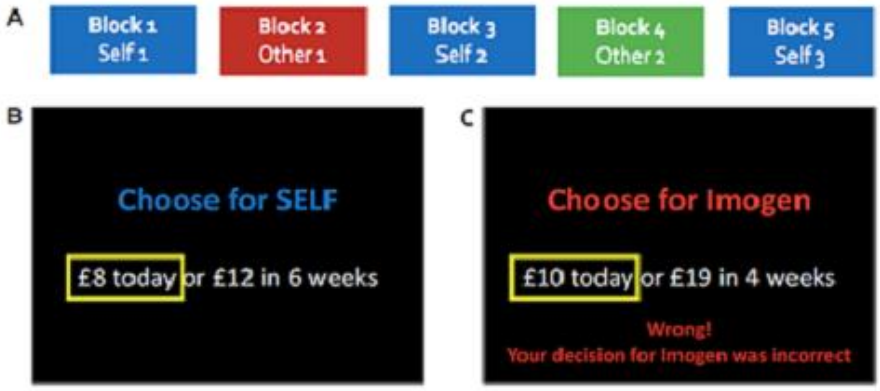


Figure 1. Design. A) Shows the order and type of the five blocks for each task. The Self, Other1 and Other2 colours were randomised across participants (RGB colours). B) Shows an example experimental screen for Self-trials, participants saw the type of trial (Self or Other), indicated by the name in colour, along with the options. The yellow box appeared around the participant's selected choice, following key press. C) Shows an example experimental screen for Other-trials, which follow the same format as Self-trials, with additional feedback. This feedback was presented on screen 500 ms after the yellow box appeared and was displayed for 1000 ms before the screen cleared for the next trial.

Study 2 and 3

Participants completed a TD contagion task which followed the same format as the Study 1 task, and the Garvert et al. (2015) task, with five trial blocks (Self1, Other1, Self2, Other2, and Self3), and two simulated agents. The number of trials in each block was reduced in this task from 50 to 30, and participants received a self-timed break after 15 trials. Participants also completed a PD contagion task as part of a separate project see **chapter 4** for analyses with this data). The order of presentation (i.e., TD contagion task then PD contagion task or vice versa) was counterbalanced across participants.

Participants in Study 2 and 3 were once again informed that the choices of the two other agents were the choices of two previous participants. In fact, the names of these two agents were chosen at random from a list of popular UK names (50 of each gender, with a different random allocation for each task), and were gender-matched to the

participant. The choices of these two agents were modelled following the same procedure as Study 1.

Estimation of discount rates and simulation of the Other's choices

Study 1, 2, and 3

A log hyperbolic model was fitted to participant's choices, and the choices of the two modelled agents were also produced using this model. The subjective value of each choice was calculated on each trial according to the following equation:

(1) $$V = \frac{M}{1+kD}$$

In this equation, subjective value, and the reward magnitude are represented by V, and M respectively, *k* represents the agent's discount rate, and *D* indicates the delay period in days. The log *k* parameter values were set between -4 and 0. A log *k* value of -4 indicates that the individual is barely sensitive to delay and bases their decision only on reward magnitude. As log *k* approaches 0, individuals are more sensitive to delay and discount rewards more steeply as a function of time (Carlisi et al., 2017; Garvert et al., 2015). The value of the immediate option will henceforth be referred to as V_{SS}, and the value of the delayed option will be referred to as V_{LL}. The subjective value of V_{SS} will always correspond to the magnitude of the reward (*M*), because the delay period in days (*D*) is 0.

The following softmax function transformed the subjective values of each option into choice probabilities (i.e., P_{SS} and P_{LL}):

(2) $$P_{LL} = \frac{1}{1+e^{-\beta(V_{LL}-V_{SS})}}$$

In this equation, β is a free parameter, which characterises noise in an individual's choices. Log β parameters were set between -1 and 1. Values closer to -1 indicate larger

non-systematic deviations around the indifference point (i.e., the point at which both choice options are equally preferred).

During Other-trials, choices were modelled to reflect the preferences of an agent whose discount rate (log *k*) was plus one or minus one from the participant´s own log *k* (calculated within the experimental script, based on the Self1-block). Differences in subjective value of the two options were translated into choice probabilities using the softmax function (equation (2)) with the temperature parameter β fixed at 1. The direction (±1) of Other1 and Other2 was counterbalanced across participants within the experimental script. The agent with a larger discount rate discounts rewards more steeply than the participant and tends to prefer options that are immediately available (i.e., more impulsive), whereas the agent with the smaller discount rate waits longer for rewards than the participant (i.e., more patient).

Participants' own log *k*, and log β values were also derived from equations (1), and (2), and were updated on a trial-by-trial basis using a Bayesian model. A uniform prior was updated on each trial, and the posterior was calculated as the likelihood of an individual's choice given the parameters log *k*, and log β weighted by this prior. This posterior was then set as the prior for the next trial and was updated on a trial-by-trial basis. To avoid the influence of strong priors for the other agents on the estimation of priors for Self, priors were reset at the beginning of the Self2 and Self3 blocks. As such, the posterior from the end of the Self1-block then became the prior for the start of the Other1-block, and the posterior from the end of the Self2-block then became the prior for the start of the Other2-block. This is in line with previous studies that have shown that we use our own priors as the starting point for learning about others (Lockwood et

al., 2018; Tarantola, Kumaran, Dayan & De Martino, 2017). The log β parameter was set to start at 0.3 for Self-blocks and was fixed at 1 for all Other-blocks.

Contagion when accounting for uncertainty

Study 1, 2, and 3

A Bayesian belief update measure was used to calculate the change in participants' log *k* values after learning about the preferences of the two other agents (i.e., contagion). The change in priors was calculated using Kullback-Leibler divergence (D_{KL}), which quantifies the divergence in the distribution of two data sets (Kullback & Leibler, 1951), and accounts for both the change in the overall peak of this distribution, and the precision of the distribution. The precision reflects an individual's confidence (or certainty) in their belief (Moutoussis et al., 2016). In the studies presented here, D_{KL} quantifies the divergence in the posterior distribution between the end trials of two blocks, following trial-by-trial update of the prior. See **Figure 2** for these posterior distributions plotted for an example participant. D_{KL} was normalised for each analysis, such that changes in log *k* in the same direction as the other agent (e.g., the participant's log *k* became more positive (i.e., impulsive) after making choices on behalf of a more impulsive agent) resulted in positive D_{KL} values, and changes in log *k* in the opposite direction resulted in negative D_{KL} values. D_{KL} was also sorted and analysed separately for impulsive and patient others (i.e., D_{KL} for a more positive/impulsive or a more negative/patient agent). Shift variables (as per Garvert et al., 2015) are outlined and included in the **Supplemental Information** for comparison.

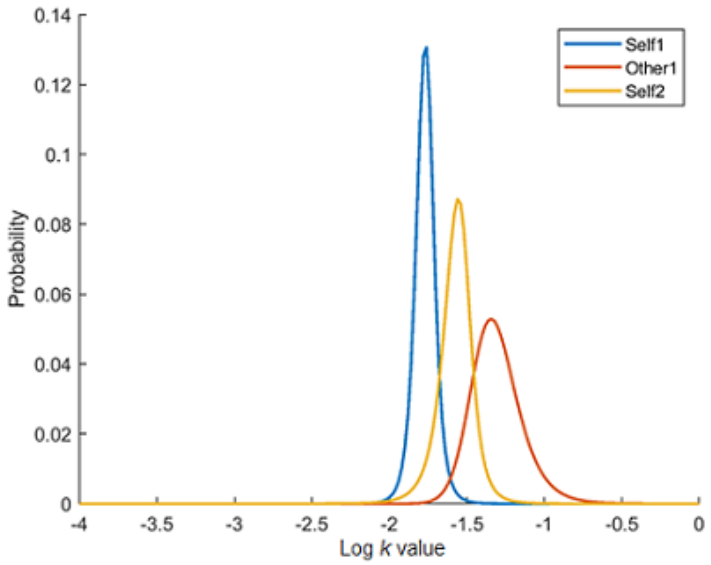


Figure 2. Example participant behaviour: These three Gaussian distribution curves show the posterior distribution at the end of the Self1-, Other1-, and Self2-blocks for an example participant (Study 2 sample). At Self2, the posterior has shifted towards the posterior for Other1 (i.e., contagion).

Individual differences in social cognition self-report measures

Study 1, 2, and 3

Participants also completed a block of six individual differences questionnaire measures. In Study 1, these were filled out following completion of the TD task. In studies 2, and 3, these were filled out in between the two discounting tasks (i.e., TD and PD). The questionnaire block (which included the AQ; Baron-Cohen et al., 2001) was presented full screen using Qualtrics. Participants completed the questionnaires in the following order: Social Network Index (SNI; Cohen et al., 1997), Apathy Motivation Index (AMI; Ang et al., 2017), Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers et al., 2011), Twenty-Item Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994), Short Form Self-Report Psychopathy scale (SRP-SF; Gordts et al., 2017), and AQ.

AQ data are analysed in relation to contagion variables in this paper, and data from the remaining questionnaires are not included, as this data was collected for a collaboration project. A Likert scoring method (scored 1-4, as introduced by Austin,

2005) is used here for scoring the AQ, instead of the original binary scoring method (Baron-Cohen et al., 2001). Higher internal consistency and greater test-retest reliability has been found for the four-point Likert scoring method compared to binary scoring using a sample of NT participants (Stevenson & Hart, 2017). As in the original measure, higher scores indicate a greater number of autistic traits. In Studies 1, and 2, individual differences are indexed using the subscale scores on the AQ. Likert-scored AQ scores for the final samples (following the exclusions outlined at the end of the **Methods** section) for all three study groups are presented in **Table 1**, and binary-scored AQ scores are presented in **Supplemental Table 1**.

Awareness of manipulation and behavioural change

Study 2, and 3

Two questions (with *"yes"* or *"no"* responses) were added to Studies 2 and 3 to determine whether participants believed that the other agents were real participants (i.e., were unaware of the manipulation; *"Did you believe that the other two players were real participants?"*), and whether they were aware of their behaviour changing (*"Did you notice yourself changing your choices throughout the experiment?"*). These questions were asked following completion of both discounting tasks and the questionnaire block. If answering "*no*" to the first question, participants were prompted to provide a reason.

Participant exclusions

Study 1

One outlier was excluded from this sample (i.e., ± 3 SDs of D_{KL}). Four participants in the final sample with log *k* values close to -4 also had two other agents

with behaviour (log *k*) in the same direction. For these participants, data were analysed for the other agent (Other1 or Other2) with the greatest distance between the participant's own discount rate, and the model discount rate. Data from 48 participants were entered into the final analysis.

Study 2

Two outliers (using the same criteria as Study 1) were excluded from this sample. From the final sample, three participants had two agents with behaviour in the same direction. As in Study 1, data were analysed for the other agent with the greatest distance between the participant's own discount rate, and the model discount rate. Data from 98 participants were entered into the final analysis.

Study 3

Two outliers were excluded, and data were analysed for only one agent for two participants in the final sample. Data from 12 participants were included in the final analysis.

Statistical analyses

Analyses for all three studies are run in log space. All analyses were run using Jamovi (version 1.1.9), JASP (version 0.13.0.0), and R (version 3.5.2) in RStudio (version 1.1.463). G*Power (version 3.1) was used for power calculations, and figures were produced in R. Bayes factors (calculated using Jamovi jsq package presets) are reported for all analyses and are interpreted in line with the jsq package criteria. Bayes factors allow the user to quantify the evidence in favour of one hypothesis over another (e.g., the hypothesis of interest versus the null), given the data (Kass & Raftery, 1995; Quintana & Williams, 2018). In the analyses presented below, BF_{10} is used for

significant analyses and denotes the likelihood that the hypothesis of interest is correct. BF_{01} is used for non-significant analyses and denotes the likelihood that the null hypothesis is correct.

Study 1, and 2

To determine whether a significant contagion effect was present, separate one-sample t-tests (or Wilcoxon signed-rank tests for non-normally distributed data) were run on D_{KL} data (separately for more patient and more impulsive agents). Robust repeated-measures t-tests were then run on D_{KL} data to assess the effect of direction (i.e., patient versus impulsive other) on contagion. Repeated-measures t-tests were also used to assess the effect of direction on accuracy in these samples, and correlation analyses were performed to determine whether a relationship could be observed between accuracy and D_{KL}. Exploratory independent-measures t-tests were conducted to determine gender effects on TD task parameters (i.e., log k and log β). To assess the association between autistic traits (as measured by the AQ) and contagion in the two NT samples, data were collapsed across these samples, and entered into two regression analyses.

Further analyses (robust independent-measures t-tests) were conducted on the Study 2 data only, to determine whether awareness of manipulation and behavioural change (assessed by the final two questions included in this study) were related to D_{KL}.

Study 3

A subset of NT participants from the Study 2 sample were selected for comparison with the Study 3 ASC sample, and the two groups were compared using Yuen's robust independent-samples t-tests on the following variables: AQ total score, AQ subscales, age, log k, log β, D_{KL}, and accuracy. The two groups were also tested for

equivalence across samples using Bayesian independent-measures t-tests for both accuracy and D_{KL} variables.

Results

Summary descriptive statistics for all following analyses are presented in **Table 2**, descriptive statistics for the AQ subscales and total scores for all three samples are included in **Table 1**. Bayes factors and interpretations are included in **Supplemental Table 2**. Analyses for the Study 1 and 2 NT samples are presented together, and analyses for the Study 3 ASC sample are presented separately.

Social contagion when accounting for uncertainty (D_{KL})

To determine whether there was a significant change in behaviour after learning the value preferences of another person (i.e., contagion), one-sample t-tests were run on the Study 1 and Study 2 sample data. If data were not normally distributed (determined by Shapiro-Wilk tests of normality, see **Supplemental Information**), Wilcoxon signed-rank tests were used. Tests were conducted separately for more impulsive agents (log *k* +1) and more patient agents (log *k* -1). See **Table 2** and **Figure 3** for descriptive statistics (mean and SD) for the D_{KL} variables for all three study samples. Contagion was significant after learning about patient others in both the Study 1 (W (47) = 1005, p < .001, d = .62, BF_{10}= 284.98) and Study 2 (t (94) = 3.60, p < .001, d = .37, BF_{10}= 41.76) samples. Contagion was also significant for impulsive others in the Study 1 sample (t (43) = 2.05, p = .046, d = .31, BF_{10}= 1.10), but not in the Study 2 sample (W (97) = 2707, p = .319, d = .13, BF_{01}= 4.10).

Table 2. Descriptive statistics (mean (SD±)) for the final samples (following exclusions) for both NT study samples (Study 1 and Study 2) and the ASC study sample (Study 3).

	Study		
	Study 1 (N=48)	Study 2 (N=98)	Study 3 (N=12)
Age	23.81 (3.86)	21.41 (2.09)	22.33 (3.63)
Gender	23 female	59 female	8 female
Log k	-2.17 (.84)	-1.68 (.67)	-2.28 (.78)
Log β	-.13 (.27)	-.19 (.35)	-.08 (.52)
D_{KL}			
Impulsive (log k +1)	.96 (3.08)	.31 (2.39)	1.50 (3.31)
Patient (log k –1)	2.14 (3.44)	.90 (2.45)	1.36 (3.57)
Percent accuracy			
Impulsive (log k +1)	80.18 (5.65)	74.46 (11.18)	75.33 (8.92)
Patient (log k -1)	83.79 (6.05)	78.60 (10.30)	79.72 (12.75)

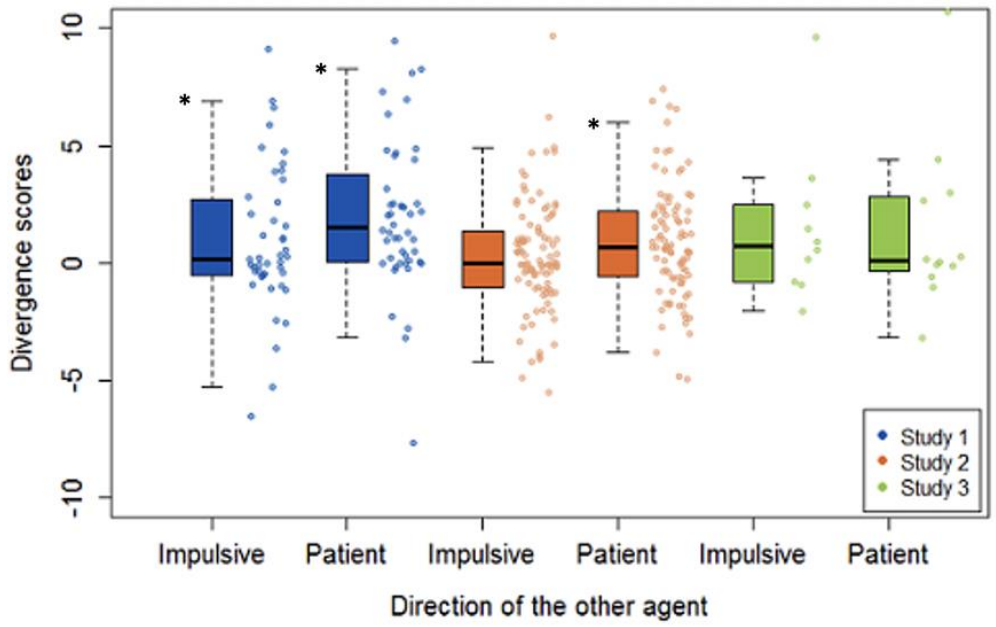


Figure 3. Group comparisons of D_{KL} scores: Descriptive statistics for all divergence score (D_{KL}) variables for the two NT samples (Study 1 and Study 2), and the ASC sample (Study 3) are presented here, split for more impulsive (log k +1) and more patient (log k -1) agents. Positive values indicate a shift towards the TD preferences of the other agent (i.e., contagion), whereas negative values indicate a shift away. One-sample t-tests (or Wilcoxon signed-rank tests where there were non-normal distributions) were conducted on Study 1 and Study 2 data, and significant results ($p < .05$) are indicated with an asterisk.

To determine whether there was an effect of direction (e.g., greater shifts in behaviour after learning about a more impulsive versus a more patient other) robust repeated measures t-tests were performed on the D_{KL} values to allow for non-normally distributed data. There was a significant effect of direction in the Study 1 sample (t (27) = -2.07, p = .048, $d = .32$, BF_{10} = .91), with a stronger contagion effect for more patient others. There was no significant effect of direction on D_{KL} values in the Study 2 sample (t (56) = -1.57, p = .121, $d = .19$, BF_{01} = 2.95).

These results demonstrate a significant contagion effect in both Studies 1 and 2, indicating that participants were influenced by the choices of the other agents. Although individuals did show greater contagion when exposed to a more patient other, this effect

was only significantly greater than contagion to a more impulsive agent in the Study 1 sample, and the Bayes factor for this finding suggests that the effect is anecdotal.

Accuracy when making choices for the other agent

Repeated measures t-tests were used to determine whether there was a significant effect of direction on accuracy. In the Study 1 sample, there was a significant effect of direction on accuracy (t (43) = -2.50, p = .016, d = -.38, BF_{10} = 2.60), with greater accuracy when making choices on behalf of more patient agents (log k -1) versus more impulsive agents (log k +1). This finding was replicated in the Study 2 sample (t (94) = -2.71, p = .008, d = -.28, BF_{10} = 3.51). As in the Study 1 sample, accuracy was higher for patient agents versus impatient agents. See **Table 2** for descriptive statistics including percentage accuracy, which accounts for the different number of trials across samples (50 trials per block in Study 1, and 30 trials per block in Studies 2 and 3).

To determine whether accuracy was related to the strength of contagion, correlation analyses between D_{KL} and accuracy were run for each direction for the two NT samples. In the Study 1 sample, the correlation between D_{KL} and accuracy was not significant for impulsive (r = -.06, p = .708, BF_{01} = 4.97), or patient other agents (r = .17, p = .243, BF_{01} = 2.86). In the Study 2 sample, there was a significant positive correlation showing that stronger contagion (D_{KL}) was associated with greater accuracy for impulsive agents (r = .27, p = .007 BF_{10} = 4.56; **Figure 4**), although the correlation was not significant for more patient agents (r = .15, p = .149, BF_{01} = 2.79).

Whilst these findings indicate that NT participants were more accurate at making choices on behalf of more patient agents, versus more impulsive agents, Bayesian support for these findings was anecdotal. As the correlation between contagion (D_{KL}) and

accuracy was only significant for the more impulsive agent in the Study 2 sample (with moderate Bayesian support), these findings suggest that there is not a reliable relationship between accuracy and contagion in NT individuals.

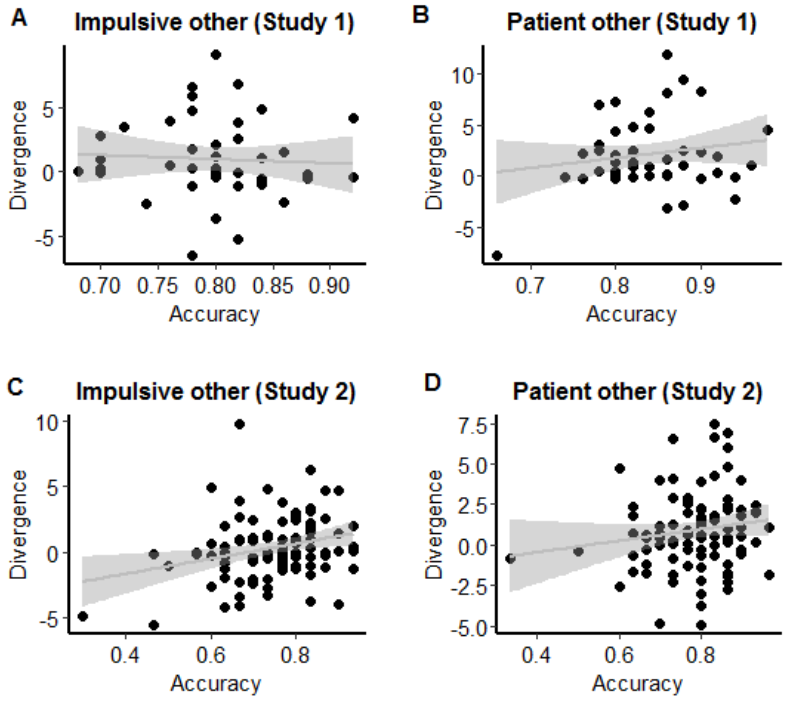


Figure 4 Correlations between percentage accuracy and D_{KL}. A) Shows the correlation for more impulsive agents in the Study 1 sample, and B) for more patient agents in the Study 1 sample. C) Shows the correlation for impulsive agents in the Study 2 sample, only this correlation was significant. D) Shows the correlation for patient agents in the Study 2 sample.

Awareness of manipulation and change, and effect on contagion

From the final sample (Study 2 participants only), a total of 87 participants completed the awareness of manipulation question (40.23% believed manipulation), and 74 participants completed the change awareness question (56.76% noticed their behaviour changing). To determine whether participant's answers to these questions were related to contagion, two robust independent measures t-tests were run, on mean D_{KL} variables.

Whilst the percentage of participants believing the manipulation is close to chance, there was no significant effect of awareness of manipulation on mean D_{KL} (t

(40.30) = 1.24, p = .223, ξ = .22, BF_{01} = 2.52). There was also no significant effect of awareness of behavioural change on mean D_{KL} (t (26.14) = 1.06, p = .298, ξ = .18, BF_{01} = 1.87). These results indicate that participants were influenced by the choices of the other to the same degree, regardless of whether they believed that the other participants were real, or whether they noticed their own behaviour changing throughout the task.

Gender differences in individual task parameters

We also conducted exploratory analyses to determine whether there were any gender differences in task parameters. Independent-measures t-tests revealed no significant gender difference in log k in the Study 1 (t (44) = -1.70, p = .096, d = -.50, BF_{01} = 1.07), or Study 2 samples (t (96) = 1.48, p = .142, d = .31, BF_{01} = 1.76). There was also no significant gender difference in log β for the Study 1 sample (t (44) = -.56, p = .580, d = -.17, BF_{01} = 3.01), although this difference was significant in the Study 2 sample (t (96) = 2.76, p = .007, d = .57, BF_{10} = 5.89), whereby females (mean = -.26, SD $\pm$ .37) made noisier choices than males (mean = -.07, SD $\pm$ .37).

Relationship between AQ subscales and contagion

To determine whether there were any significant relationships between D_{KL}, and the subscales of the AQ, data were collapsed across the two NT samples and entered into regression analyses (see **Table 1** for descriptive statistics for collapsed AQ data). Two regression analyses are conducted here, separately for more impulsive and more patient other agents. D_{KL} is entered as the outcome variable, and the predictors are a binary group variable (Study 1/Study 2), accuracy, and the AQ subscales (**Table 3**).

For the more impulsive agent, the overall model explained 0.8% of the variance in D_{KL}, and was not a significant predictor overall (F (7, 131) = .84, p = .556, BF_{01} = 184.26). For the more patient agent, the overall model explained 6.9% of the variance in D_{KL}, and was a significant predictor overall (F (7, 131) = 2.46, p = .021, BF_{10} = .40). In both models, none of the individual predictors significantly predicted D_{KL}, suggesting no relationship between AQ scores and contagion effects. This indicates that the level of autistic traits in our NT samples was not associated with contagion of TD value preferences.

Table 3. Statistics for the individual predictors (group, accuracy, and AQ subscales) entered into the regression model to predict D_{KL}.

	Impulsive			Patient		
	β	t	p	β	t	p
Study group	1.36	.85	.395	2.06	1.31	.193
Accuracy	.12	1.63	.105	.15	1.94	.055
AQ						
Social Skills	.02	.29	.776	-.07	-.91	.362
Attention Switching	-.01	-.07	.942	-.04	-.50	.616
Attention to Detail	.02	.28	.777	-.09	-1.52	.131
Communication	-.00	-.07	.948	.13	1.78	.075
Imagination	-.04	-.58	.561	.08	1.22	.226

Study 3

To allow for direct comparison between ASC and NT groups, a subset of NT participants (N = 12, 8 female, Study 2 sample) most closely matching the ages and

genders of the Study 3 sample (N = 12, 8 female) were selected. The samples were also matched on the direction of the other (i.e., if the Study 3 participant had two negative others, participants were selected which had either two negative others, or a positive and negative other, but not two positive others). If multiple participants from the NT sample matched a participant in the ASC sample, a participant ID was selected at random.

The two groups were compared using Yuen's robust independent-samples t-tests. The ASC group scored significantly higher on the majority of the AQ subscales, but there were no significant group differences in age, log *k*, log β, D_{KL}, or accuracy. See **Table 4** for descriptive statistics and t-test statistics for these analyses.

As the group differences in D_{KL} and accuracy were not significant, equivalence Bayesian independent-measures t-tests were also run on the data to determine whether the samples were significantly equivalent on these variables. Equivalence bounds were set between -.7 and .7 (determined by a power analysis for 33% power with a sample size of 12 in each group; Lakens, 2017).

Bayesian support for equivalence was strong for accuracy for both impulsive ($BF_{\in\notin}$ = 18.66), and patient ($BF_{\in\notin}$ = 16.00) agents. For D_{KL} Bayesian support for equivalence was moderate for impulsive ($BF_{\in\notin}$ = 4.96), and patient ($BF_{\in\notin}$ = 9.13) agents. See **Figure 5** for a comparison of the two samples. Together, these findings indicate that both contagion (D_{KL}) and accuracy for learning the preferences of another agent were statistically similar across the ASC and NT samples. This demonstrates that despite clear differences in social skills, autistic participants were able to learn the preferences of others and were influenced by them in the same way as NT controls.

Table 4. Descriptive statistics (mean (SD)), and comparisons (Yuen's robust independent-measures t-tests) between the ASC and matched NT sample for: age, log *k*, log β, D_{KL}, accuracy and AQ subscales.

	Group		Statistic				
	ASC	NT	t	df	*p*	ξ	*BF*
Age	22.33 (3.63)	22.42 (3.55)	.07	14.00	.948	.02	2.68
Log *k*	-2.28 (.78)	-1.57 (.81)	1.69	12.2	.116	.50	.53
Log β	-.08 (.52)	-.32 (.22)	1.39	10.74	.193	.51	1.24
D_{KL}							
Impulsive (log *k* +1)	1.50 (3.31)	.41 (.85)	.51	6.21	.627	.23	1.68
Patient (log *k* -1)	1.36 (3.57)	.45 (1.57)	.01	9.50	.992	.00	2.12
Percent accuracy							
Impulsive (log *k* +1)	75.33 (8.92)	75.28 (11.50)	.37	10.85	.722	.12	2.59
Patient (log *k* -1)	79.72 (12.75)	77.78 (8.80)	.90	13.77	.383	.22	2.50
AQ							
Social Skills	29.83 (4.13)	19.92 (3.70)	6.40	13.74	<.001*	.87	4096.01
Attention Switching	33.17 (3.66)	23.58 (3.20)	6.99	13.87	<.001*	.97	14615.95
Attention to Detail	29.42 (4.58)	24.92 (4.83)	2.10	13.98	.054	.57	.41
Communication	31.67 (3.89)	20.00 (4.31)	6.42	14.00	<.001*	.91	19376.95
Imagination	25.42 (3.83)	19.83 (3.71)	3.06	13.15	.009*	.80	22.21
Total	149.50 (14.39)	108.25 (11.96)	7.46	13.84	<.001*	.88	72616.40

Note: BF is BF_{10} for significant analyses and BF_{01} for non-significant analyses. * is for significant analyses.

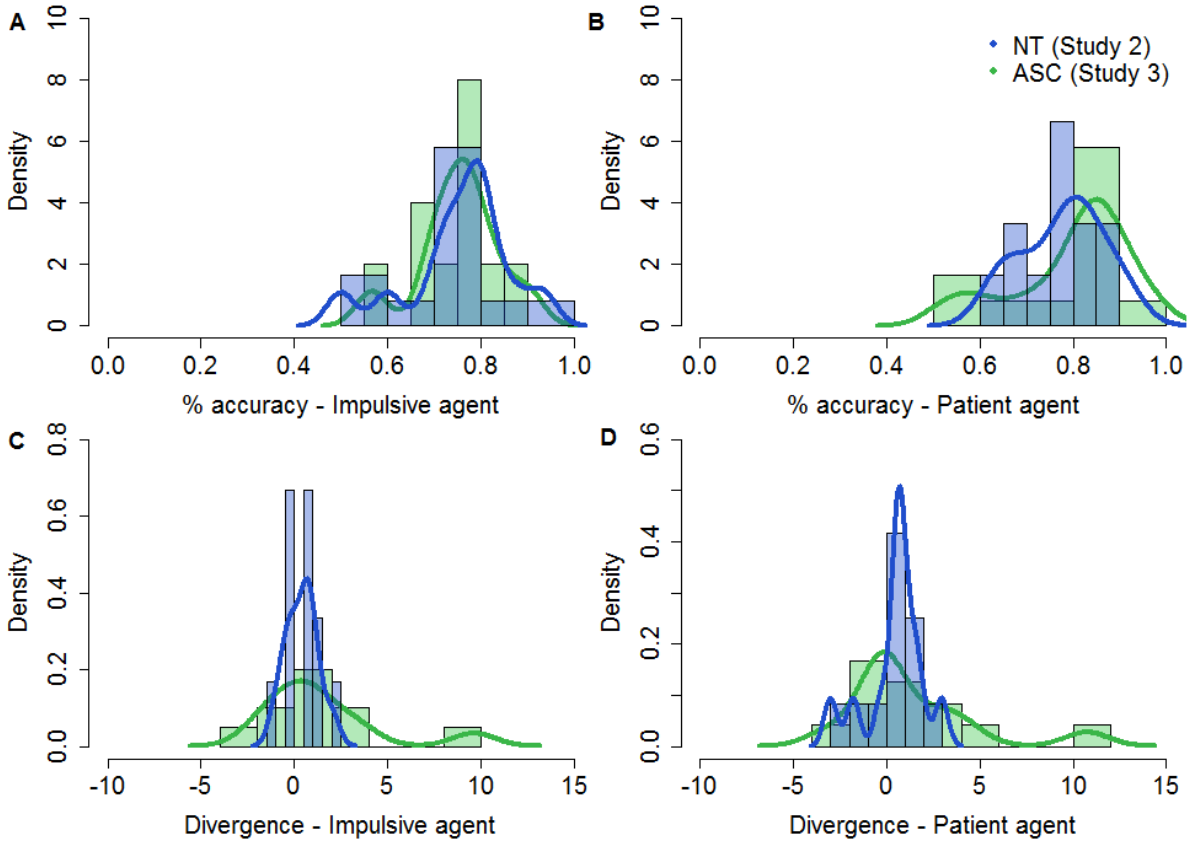


Figure 5 Histograms and density plots with comparisons between the NT subset and ASC sample. A) Shows the comparison between samples for percentage accuracy for the impulsive agent, B) shows the comparison between samples for percentage accuracy for the patient agent, C) shows the comparison between samples for D_{KL} for the impulsive agent, and D) shows the comparison between samples for D_{KL} for the patient agent.

Discussion

Across three studies, we examined contagion of value preferences on a TD task, and whether contagion correlates with autistic traits and is disrupted in autistic adults. In both of our NT samples, we found significant contagion effects (i.e., a change in value preference after exposure to another person's distinct value preference). The strength of contagion was not associated with belief in the agent, accuracy of learning about the other agent, or autistic traits. Analyses also provided support for equivalent contagion and accuracy across the ASC sample and a matched subset of NT participants. These

findings suggest that the ability to learn the value preferences of others is not different in autistic individuals and shows no relationship with autism traits in the NT population. These results add to the growing literature exploring social influence in autism and questioning the extent of social deficits in autism.

Individual differences in contagion in neurotypical samples

In line with previous research, we found that learning the different discounting preferences of another agent led to significant shifts in the participant's own discount rate in two independent NT samples (Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012). Across our two independent NT samples, contagion effects were clearly present when participants learned the preferences of more patient agents, with weaker evidence for contagion effects for more impulsive agents. This bias towards becoming more patient has also been demonstrated by Moutoussis et al. (2016) using a more advanced computational modelling approach and a similar TD contagion paradigm. However, it should be noted that bias towards becoming more patient replicates even when using a less advanced computational modelling approach that does not take into account the precision of the prior beliefs (see '**normalised shift in discount rate'** in **Supplemental analyses**). Effect sizes and Bayes Factors indicated that contagion is a strong and reproducible phenomenon; however, we were unable to associate individual differences in belief that the agent was real with the size of the contagion effect. There was anecdotal support for a relationship between how accurately someone can predict the choices of another agent and contagion in one of our NT samples. These findings highlight the need for further research investigating variability in the strength of contagion and why some individuals are more influenced by others.

Contagion effects in autism

Contrary to predictions, subscale scores on the AQ were not significantly related to contagion in the two NT samples. Whilst we did find group differences in AQ scores (with higher scores for the Study 1 sample), we collapsed across samples for our regression analysis predicting contagion. Furthermore, using equivalence tests, we found equivalent contagion in our ASC sample (Study 3) and a subset of NT participants (Study 2 sample). These findings challenge the extent to which social influence differs between autistic and NT populations.

Previous research into social influence in autism has focused on conformity (an explicit social influence effect; Asch, 1956), and the findings have been varied (Lazzaro et al., 2018; Van Hoorn et al., 2017; Yafai et al., 2014). Lower rates of conformity have been found in autistic children compared to typically developing controls for a line judgement task administered before and after participants received incorrect social information from the experimenter (i.e., *"Most people think…"*; Yafai et al., 2014). In a public goods game with peers, decreased conformity has been found for antisocial peer influence in adolescents with high levels of autism traits (across NT and autistic samples), whereas comparable levels of prosocial peer influence were found across groups (Van Hoorn et al., 2017). Autistic adults (as in the present study) have also been shown to be as susceptible to conformity as NT controls in tests of word memory (Lazzaro et al., 2018). Here, we use a different form of social influence (i.e., contagion) in combination with a distinct value-based decision-making paradigm. In line with Lazzaro et al. (2018), and Bowler and Worley (1994), we found no significant differences in contagion between autistic and NT samples. In addition, we also fail to find a relationship between any of the AQ subscales and the strength of contagion in the two

NT samples. Finally, our equivalence tests provide significant support for similarity between the ASC sample and a subset of NT participants, thus not relying on the absence of an effect to draw our conclusions. Together, these findings support the notion that social influence (both conformity and contagion) is unrelated to ASC.

Learning about others in ASC

Our results also suggest that autistic individuals learnt the value preferences of other agents just as accurately as NT participants. Difficulties learning the distinct perspectives of another agent, often referred to as Theory of Mind (ToM), has characterised social impairments in autism (Baron-Cohen et al., 1985; Premack & Woodruff, 1978). Whilst multiple previous studies have suggested that autistic individuals struggle to correctly attribute mental states to others (Baron-Cohen et al., 1985), an increasing number of studies have shown that the performance of autistic individuals depends on multiple factors, including: comorbidities, task framing, and motivation (Frith & Happé, 1994; Hamilton, 2009; Keysar et al., 2003; Oakley et al., 2016; Peterson et al., 2013; Senju et al., 2009; Shamay-Tsoory et al., 2019). Our results provide further evidence that autistic individuals are able to learn the distinct preferences of another person. In particular, they suggest that at least when feedback is explicit, social impairments are not observed. It could be that without explicit feedback, autistic individuals would show impairment on the task. Indeed, research has found that social information is less salient for autistic individuals than NT individuals (Freeth et al., 2011), and predictive coding models of autism have suggested that autistic individuals generate less precise internal prior beliefs about the world and rely more heavily on external sources of information (Karvelis et al., 2018; Lawson et al., 2017; Pellicano &

Burr, 2012; Van De Cruys et al., 2014). The implicit nature of social cues makes it more difficult to generate robust internal priors about others' beliefs, as these external sources of information are inherently noisy and less precise. Our results suggest that by using more explicit social feedback, deficits in learning about others were extinguished such that accuracy was comparable between ASC and NT samples. Future research could manipulate the consistency of the other agents to account for the greater noise seen in social behaviour and assess learning for ASC and NT samples.

Caveats and conclusions

Taken together, these findings demonstrate that learning the value preferences of other agents robustly shifts the value preferences of NT participants (i.e., contagion). In contrast to our hypothesis, there was no reliable evidence of a relationship between autistic traits and contagion in NT samples, and both contagion and social learning was similar across ASC and NT samples.

However, it is important to highlight two potential limitations. First, studies exploring a range of behavioural measures have shown that it is difficult to link individual differences in questionnaires with task data due to differences in variability (Frey et al., 2017; Hedge et al., 2017; Palminteri & Chevallier, 2018; Pedroni et al., 2017). Behavioural tasks (such as discounting tasks) produce replicable results because between-subject variability is low, which results in low reliability for measuring individual differences (Hedge et al., 2017). Nevertheless, similar research by Molleman et al. (2019) and Reiter et al. (2019) has found that the strength of contagion correlated with individual differences in social conformity and social integration values respectively. Whilst variability in our collapsed sample was similar to that of previous

research (e.g., Goris et al., 2019, who reported an SD of 14.47, compared to 15.34 in our collapsed sample), we did not compare differences in contagion between high and low AQ scorers in our study. Future research could seek to recruit participants both with high and low AQ scores, in order to determine whether group differences exist here. Much of the data in our two NT samples was also collected from a university population, and further research should explore whether relationships between contagion and individual differences in AQ scores become clearer if all participants are recruited from the general population.

Second, it is possible that the lack of a significant group difference (i.e., ASC versus NT subset), and the significant support for similarity between groups reflects under sampling of the autistic population (N = 12). However, power calculations indicated that only four participants would be required to demonstrate a contagion effect (with 80% power) based on previous studies, and thus our study was well-powered. Nevertheless, future studies could seek to replicate the effects in a larger sample of autistic participants, given the very large effect sizes shown in previous contagion studies and the significant similarity between ASC and NT samples. We believe an effect of contagion of value preferences in autism suggest a social context where Bayesian beliefs are accurately formed, and our findings suggest that exploring social influence using a Bayesian approach is a promising avenue for future research into autism.

References

Ang, Y. S., Lockwood, P., Apps, M. A. J., Muhammed, K., & Husain, M. (2017). Distinct subtypes of apathy revealed by the Apathy Motivation Index. *PLOS ONE, 12*(1), 1-15.

Antrop, I., Stock, P., Verté, S., Wiersema, J. R., Baeyens D., & Roeyers, H. (2006). ADHD and delay aversion: The influence of non-temporal stimulation on choice for delayed rewards. *Journal of Child Psychiatry, 47*(11), 1152-1158.

Apps, M. A. J., & Ramnani, N. (2017). Contributions of the medial prefrontal cortex to social influence in economic decision-making. *Cerebral Cortex, 29*(9), 4635-4648.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1-70.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences, 38*(2), 451-460.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale: Item selection and cross validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23-32.

Balsters, J. H., Apps, M. A. J., Bolis, D., Lehner, R., Gallagher, L., & Wenderoth, N. (2017). Disrupted prediction errors index deficits in autism spectrum disorder. *Brain, 140*(1), 235-246.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*(7219), 245-249.

Bennett, D. (2015). The neural mechanisms of Bayesian belief updating. *The Journal of Neuroscience, 35*(50), 16300-16302.

Bowler, D. M., & Worley, K. (1994). Susceptibility to social influence in adults with Asperger's syndrome: A research note. *The Journal of Child Psychology and Psychiatry, 35*(4), 689-697.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology, 20*(13), 1165-1170.

Carlisi, C. O., Norman, L., Murphy, C. M. Christakou, A., Chantiluke, K., Giampietro, V., … Rubia, K. (2017). Comparison of neural substrates of temporal discounting between youth with autism spectrum disorder and with obsessive-compulsive disorder. *Psychological Medicine, 47*(14), 2513-2527.

Chantiluke, K., Christakou, A., Murphy, C. M., Giampietro, V., Daly, E. M., Ecker, C., … Rubia, K. (2014). Disorder-specific functional abnormalities during temporal discounting in youth with attention deficit hyperactivity disorder (ADHD), autism and comorbid ADHD and autism. *Psychiatry Research, 223*(2), 113-120.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591-621.

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *The Journal of the American Medical Association, 277*(24), 1940-1944.

D'Cruz, A. M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., & Sweeney, J. A. (2013). Reduced behavioural flexibility in autism spectrum disorders. *Neuropsychology, 27*(2), 152-160.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgements according to its reliability. *The Journal of Neuroscience, 37*(25), 6066-6074.

Demurie, E., Roeyers, H., Baeyens, D., & Sonuga-Barke, E. (2012). Temporal discounting of monetary rewards in children and adolescents with ADHD and autism spectrum disorders. *Developmental Science*, *15*(6), 791-800.

Freeth, M., Ropar, D., Mitchell, P., Chapman, P., & Loher, S. (2011). How adolescents with ASD process social information in complex scenes. Combining evidence from eye movements and verbal descriptions. *Journal of Autism and Developmental Disorders, 41*(3), 364-371.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, *3*(10), 1–13.

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind". *Cognition, 50*, 115-132.

Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J. & Dolan, R. J. (2015). Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron, 85*(2), 418-428.

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & Rossi, G. (2017). Validity of the Self-Report Psychopathy scales (SRP-III full and short versions) in a community sample. *Assessment, 24*(3), 308-325.

Goris, J., Brass, M. Cambier, C., Delplanque, J., Wiersema, J. R., & Braem, S. (2019). The relation between preference for predictability and autistic traits. *Autism Research, 13*(7), 1144-1154.

Hamilton, A. F. d. C. (2009). Research review: Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*(8), 881-892.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavioural Research, 50,* 1166–1186.

Izuma, K. (2013). The neural basis of social influence and attitude change. *Current Opinion in Neurobiology, 23*(3), 456-462.

Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriè, O. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife, 7,* 1-29.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25-41.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86.

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analysis. *Social Psychological and Personality Science, 8*(4), 355-362.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience, 20*(9), 1293-1299.

Lazzaro, S. C., Weidinger, L., Cooper, R. A., Baron-Cohen, S., Moutsiana, C., & Sharot, T. (2018). Social conformity in autism. *Journal of Autism and Developmental Disorders*, *49*, 1304-1315.

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications*, *9*(1), 4747.

Meyer, M. L., Zhao, Z., & Tamir, D. I. (2019). Simulating other people changes the self. *Journal of Experimental Psychology: General, 148*(11), 1898-1913.

Molleman, L., Kurvers, R. H. J. M., & van den Bos, W. (2019). Unleashing the BEAST: A brief measure of human social information use. *Evolution and Human Behavior*, *40*(5), 492-499.

Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of inter-temporal preferences. *PLoS Computational Biology*, *12*(7), 1-17.

Mussey, J. L., Travers. B. G., Klinger, L. G., & Klinger, M. R. (2015). Decision-making skills in ASD. Performance on the Iowa Gambling Task, *Autism Research*, *8*(1), 105-114.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114-1121.

Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology, 125*(6), 818-823.

Palminteri, S., & Chevallier, C. (2018). Can we infer inter-individual differences in risk-taking from behavioral tasks? *Frontiers in Psychology, 9*(2307), 1–5.

Park, S. A., Goïame, S., O'Connor, D. A., & Dreher, J. C. (2017). Integration of individual and social information for decision-making in groups of different sizes. *PLoS Biology, 15*(6), 3-28.

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour, 1*(11), 803–809.

Pellicano, E., & Burr, D. (2012). When the world becomes "too real". A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504-510.

Pellicano, E., Smith, A. D., Cristino, F., Hood, B. M., Briscoe, J., & Gilchrist, I. D. (2011). Children with autism are neither systematic nor optimal foragers. *PNAS, 108*(1), 421-426.

Perreault, C., Moya, C., & Boyd, R. (2012). A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior, 33*, 449-459.

Peterson, C. C., Slaughter, V., Peterson, J., & Premack, D. (2013). Children with autism can track others' beliefs in a competitive game. *Developmental Science, 16*(3), 443-450.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515-526.

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry, 18*(1), 178.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences 13*(10), 420-428.

Reiter, A. M. F., Suzuki, S., O'Doherty, J. P., Li, S. C., & Eppinger, B. (2019). Risk contagion affects learning and decision-making in adolescents. *Journal of Experimental Psychology: General, 148*(9), 1494-1504.

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment, 93*(1), 84-95.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883-885.

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., & Gvirts, H. Z. (2019). Herding brains: A core neural mechanism for social alignment. *Trends in Cognitive Sciences, 23*(3), 174-186.

Stevenson, J. L., & Hart, K. R. (2017). Psychometric properties of the autism spectrum quotient for assessing low and high levels of autistic traits in college students. *Journal of Autism and Developmental Disorders, 47*(6), 1838-1853.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., … Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron, 74*, 1125-1137.

Suzuki, S., Jensen, E. L. S., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioural contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *PNAS, 113*(14), 3755-3760.

Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications, 8*(817), 1-14.

Van De Cruys, S., Evers, K., Van Der Hallen, R., Van Eylen, L., Boets, B., De-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121,* 649-675.

Van Hoorn, J., Van Dijk, E., Crone, E. A., Stockmann, L., & Rieffe, C. (2017). Peers influence prosocial behavior in adolescent males with autism spectrum disorders. *Journal of Autism and Developmental Disorder, 47*(7), 2225-2237.

Warnell, K. R., Maniscalco, S., Baker S., Yi, R., & Redcay, E. (2019). Social and delay discounting in autism spectrum disorder. *Autism Research, 12*(6), 870-877.

Wheeler, L. (1966). Toward a theory of behavioral contagion. *Psychological Review, 73*(2), 179-192.

Wu, H. C., White, S., Rees, G., & Burgess, P. W. (2018). Executive function in high-functioning autism: Decision-making consistency as a characteristic gambling behaviour. *Cortex, 107,* 21-36.

Yafai, A. F., Verrier, D., & Reidy, L. (2014). Social conformity and autism spectrum disorder: A child-friendly take on a classic study. *Autism, 18*(8), 1007-1013.

Supplemental Information

Supplemental methods

Generation of choice pairs

Study 1

To ensure accurate estimations of participants' discount rates, choice pairs for all Self-trials were generated according to two methods: generative and adaptive. These were alternated across trials (as per Garvert et al., 2015). The generative method involved generating all possible pairs of amounts and delays. For each Self-block, 25 trials were selected that best matched the indifference points of 25 hypothetical participants with log *k* values evenly distributed between -4 and 0 (Garvert et al., 2015; Nicolle et al., 2012). All 50 Other-trial choice pairs were generated according to the same method, so that the options presented to participants best matched the indifference points of 50 hypothetical subjects with evenly distributed log *k* values.

The remaining 25 trials for each Self-block were generated according to an adaptive method, which relied on a Bayesian framework to produce precise estimations of discounting parameters. This method has been shown to produce more reliable estimations of *k* value with fewer trials required (Vincent & Rainforth, 2017). The participant's initial prior belief about log *k* (at the start of the first block) was set to be normally distributed, with a mean of -2 and a standard deviation of 1. After each decision, Bayes rule was used to form the posterior by updating this prior. Choice pairs were generated which probed participant's indifference points, in order to refine the prior.

Study 2, and 3

As in Study 1, choice pairs for all Self-trials were generated according to the same adaptive and generative methods, alternated across trials. For each Self-block, the generative method was used to generate all possible pairs of amounts and delays, and 15 trials were selected that best matched the indifference points of 15 hypothetical participants with log *k* values evenly distributed between -4 and 0. All Other-trial choice pairs (30 per block) were generated according to the same method (i.e., for each block, choices best matched the indifference points of 30 hypothetical subjects with evenly distributed log *k* values). The remaining 15 trials in each Self-block were generated according to the adaptive method, in the same format as Study 1.

Supplemental analyses

Distribution of contagion variables

In the Study 1 sample, D_{KL} values for impulsive others were normally distributed (W = .95, p = .079), and D_{KL} values for patient others were not normally distributed (W = .94, p = .013). In the Study 2 sample, D_{KL} values for impulsive others were not normally distributed (W = .96, p = .009), and D_{KL} values were normally distributed (W = .98, p = .193).

Shift calculation

Previous research has included analyses using shift variables (Garvert et al., 2015), and this is also calculated here for comparison. Normalised shift was calculated for both NT samples (Study 1 and Study 2) included in this paper (i.e. shifts in the same direction as the other agent are positive, and shifts in the opposite direction to the other agent are negative). Basic (Self after Other - Self before Other) and distance-controlled

(as per Garvert et al., 2015) shift variables are included here. Distance controlled shift was calculated using the following equations:

(1) $\text{Shift1} = \frac{Self2_{logk} - Self1_{logk}}{Other1_{logk} - Self1_{logk}}$

(2) $\text{Shift2} = \frac{Self3_{logk} - Self2_{logk}}{Other2_{logk} - Self2_{logk}}$

The log *k* value from the last trial of each block was used here, as discounting parameters are stable by this point. The log *k* values calculated for the Other blocks, were calculated from participants' choices on behalf of the other agent. Supplemental Equation (1) shows the calculation of shift at Self2 away from Self1 and towards Other1. Supplemental Equation (2) shows this shift at Self3, away from Self2 and towards Other2. Shift variables were sorted into shift for impatient (i.e., more positive) and patient (i.e., more negative) other agents (based on the direction of the model, i.e., ±1 of the participant's log *k* value).

To determine whether basic and distance-controlled normalised and absolute shifts in log *k* were significantly greater than zero, separate one-sample t-tests were run for more impulsive (log *k* +1) and more patient (log *k* -1) other agents. If data were not normally distributed (determined by Shapiro-Wilk tests of normality), Wilcoxon signed-rank tests were used.

Normalised shift in discount rate

In both NT samples, basic and distance-controlled normalised shift was significantly greater than zero for patient agents only. See **Supplemental Figure 1** for descriptive statistics for all shift variables.

Study 1

Basic normalised shift in log *k* towards the more impulsive other was normally distributed (W = .96, p = .141), and was not significantly greater than zero (t (43) = 1.93, p = .060, d = .29, BF_{01} = 1.12). Basic shift towards the more patient other was not normally distributed (W = .93, p = .010) and was significant (W (47) = 1010.00, p < .001, d = .68, BF_{10} = 1007.63). Distance-controlled shift was not normally distributed (W = .72, p < .001) and was not significantly greater than zero (W (43) = 638.00, p = .097, d = .18, BF_{01} = 3.15) for the impulsive agent. For the patient agent, distance-controlled shift was also not normally distributed (W = .29, p < .001), but was significantly greater than zero (W (47) = 930.00, p < .001, d = -.09, BF_{10} = .19).

Study 2

In the Study 2 sample, basic shift was not normally distributed (W = .94, p < .001), and was not significantly greater than zero for the impulsive agent (W (97) = 2683.00, p = .362, d = .04, BF_{01} = 8.18). Basic shift for the more patient agent was normally distributed (W = .98, p = .294), and was significantly greater than zero (t (94) = 4.31, p < .001, d = .44, BF_{10} = 438.35). Distance-controlled shift was not normally distributed for the more impulsive agent in this sample (W = .48, p < .001), and was also not significantly greater than zero (W (97) = 2670.00, p = .387, d = .02, BF_{01} = 8.70). Distance-controlled shift was also not normally distributed for the more patient other agent in this sample (W = .81, p < .001), but was significantly greater than zero (W (94) = 3308.00, p < .001, d = .31, BF_{10} = 8.15).

Contagion when accounting for uncertainty (D_{KL}) versus shift

Bayes factors were compared for normalised distance-controlled shift (as per Garvert et al., 2015), which controls for the distance between Self and other in the

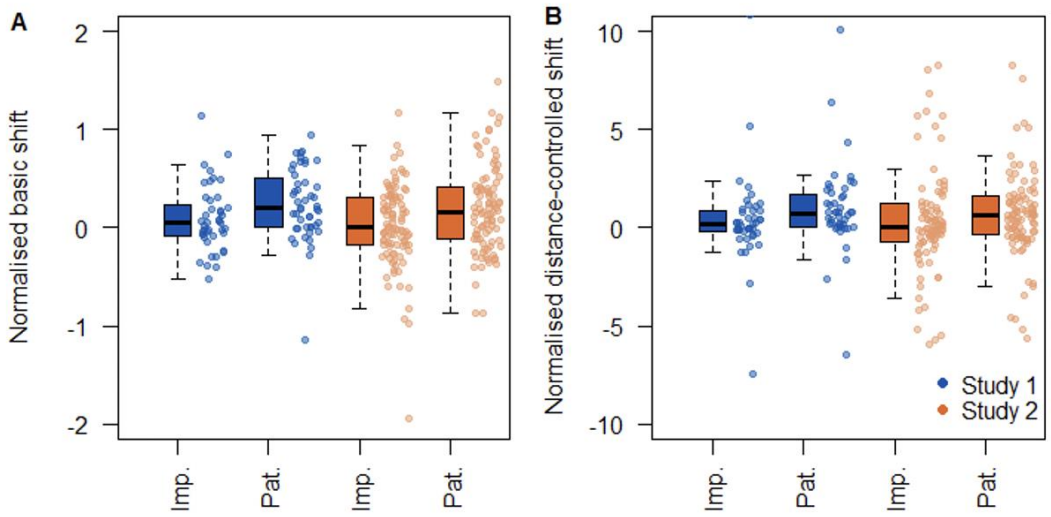
calculation, and D_{KL}, which controls for the participants' certainty in their belief in both blocks. This was to determine which variable resulted in the largest effect (Bayes factors in support of an effect (BF_{10}) were used here for comparison purposes). In both NT samples, an effect of contagion was more likely when using the normalised D_{KL} measure versus distance-controlled normalised shift.

Study 1

For the impulsive agent in this sample, an effect of contagion was 3.44 times more likely when using D_{KL} to measure contagion (BF_{10} = 1.10) than when using the shift variable (BF_{10} = .32). Contagion was also greater (x 1499.90) when using D_{KL} (BF_{10} = 284.98) versus shift (BF_{10} = .19) for the patient agent in this sample.

Study 2

In this sample, an effect of contagion was twice as likely when using the D_{KL} variable (BF_{10} = .24) to measure contagion, versus shift (BF_{10} = .12) for the impulsive agent, and 5.12 times more likely when using D_{KL} (BF_{10} = 41.76) versus shift (BF_{10} = 8.15) for the patient agent.



Supplemental Figure 1. Descriptive statistics for all shift variables for the two NT samples (Study 1 and Study 2) are presented here, split for more impulsive (i.e., Imp., positive, log *k* +1) and more negative (i.e., Pat., negative, log *k* -1) agents. A positive value indicates a shift in value preference towards the other agent (for example, becoming more impulsive after being exposed to an impulsive other) and a negative value indicates a shift in behaviour away from the other agent (for example, becoming more patient after learning the preferences of a more impulsive other). A) Shows normalised basic shift variables, and B) shows normalised distance-controlled variables.

Supplemental Table 1. Descriptive statistics for AQ subscale scores and total scores for both NT samples (Study 1 and Study 2), and the ASC sample (Study 3) using binary scoring (Baron-Cohen et al., 2001), following exclusions.

	Study 1			Study 2			Study 3		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
AQ									
Social Skills	5.20	1.76	1-9	2.54	1.97	0-8	6.92	1.62	4-9
Attention Switching	5.09	1.84	2-9	4.67	2.08	1-10	8.75	1.14	7-10
Attention to Detail	4.83	1.60	2-9	5.21	2.23	0-10	7.17	1.99	4-10
Communication	4.74	1.45	2-7	2.23	1.94	0-8	8.08	1.44	5-10
Imagination	5.07	1.70	1-8	2.32	1.82	0-9	5.50	1.68	3-9
TOTAL	24.91	3.85	19-36	16.97	6.19	5-36	36.42	5.05	27-43

Supplemental Table 2. Presented here are the Bayes factors and interpretations for all analyses with the Study 1 (NT), Study 2 (NT), and Study 3 (ASC) samples and the comparison samples (NT subsample and ASC sample) in the order presented in the paper (main paper, followed by Supplemental analyses) to allow for direct comparison.

	Study			
	Study 1 (NT)	**Study 2 (NT)**	**Studies 1 and 2 combined**	**Comparison (NT subset and ASC)**
One-sample t-test/Wilcoxon singed-rank test - D_{KL}, impulsive (log k +1)	$\boldsymbol{BF_{10}}$ **= 1.10** **Anecdotal**	BF_{01} = 4.10 Moderate	-	-
One-sample t-test/Wilcoxon singed-rank test - D_{KL}, patient (log k -1)	$\boldsymbol{BF_{10}}$ **= 284.98** **Extreme**	$\boldsymbol{BF_{10}}$ **= 41.76** **Strong**	-	-
Robust repeated-measures t-test – Effect of direction (+/- log k) on D_{KL}	$\boldsymbol{BF_{10}}$ **= .91**	BF_{01} = 2.95 Anecdotal	-	-

	Anecdotal support for H0			
Repeated-measures t-test – Effect of direction (+/- log *k*) on percentage accuracy	**BF_{10} = 2.60 Anecdotal**	**BF_{10} = 3.51 Moderate**	-	-
Correlation - D_{KL} and accuracy, impulsive (log *k* +1)	BF_{01} = 4.97 Moderate	**BF_{10} = 4.56 Moderate**	-	-
Correlation - D_{KL} and accuracy, patient (log *k* -1)	BF_{01} = 2.86 Anecdotal	BF_{01} = 2.79 Anecdotal	-	-
Independent-measures t-test – Effect of belief in manipulation on mean D_{KL}	-	BF_{01} = 2.52 Anecdotal	-	-
Independent-measures t-test – Effect of awareness of manipulation on mean D_{KL}	-	BF_{01} = 1.87 Anecdotal	-	-
Independent-measures t-test – Gender difference in log *k*	BF_{01} = 1.07 Anecdotal	BF_{01} = 1.76 Anecdotal	-	-
Independent-measures t-test – Gender difference in log β	BF_{01} = 3.01 Moderate	**BF_{10} = 5.89 Moderate**	-	-
Regression – Group, accuracy, and AQ subscales predicting D_{KL}, impulsive (log *k* +1)	-	-	BF_{01} = 184.26 Extreme	-
Regression – Group, accuracy, and AQ subscales predicting D_{KL}, patient (log *k* -1)	-	-	**BF_{10} = .40 Anecdotal support for H0**	-
Yuen's robust Independent-samples t-test – Group difference in log *k*	-	-	-	BF_{01} = 2.68 Anecdotal

Yuen's robust Independent-samples t-test – Group difference in log β	-	-	-	BF_{01} = .53 Anecdotal support for H0
Yuen's robust Independent-samples t-test – Group difference in D_{KL}, impulsive (log k +1)	-	-	-	BF_{01} = 1.24 Anecdotal
Yuen's robust Independent-samples t-test – Group difference in D_{KL}, patient (log k -1)	-	-	-	BF_{01} = 1.68 Anecdotal
Yuen's robust Independent-samples t-test – Group difference in % accuracy, impulsive (log k +1)	-	-	-	BF_{01} = 2.59 Anecdotal
Yuen's robust Independent-samples t-test – Group difference in % accuracy, patient (log k -1)	-	-	-	BF_{01} = 2.50 Anecdotal
Yuen's robust Independent-samples t-test – Group difference in AQ Social Skills subscale	-	-	-	**BF_{10} = 4096.01 Extreme**
Yuen's robust Independent-samples t-test – Group difference in AQ Attention Switching subscale	-	-	-	**BF_{10} = 14615.95 Extreme**
Yuen's robust Independent-samples t-test – Group difference in AQ Attention to Detail subscale	-	-	-	BF_{01} = .41 Anecdotal support for H1
Yuen's robust Independent-samples t-test – Group difference in AQ Communication subscale	-	-	-	**BF_{10} = 19376.95 Extreme**
Yuen's robust Independent-samples t-test – Group difference in AQ Imagination subscale	-	-	-	**BF_{10} = 22.21 Strong**

Yuen's robust Independent-samples t-test – Group difference in AQ total score	-	-	-	**BF_{10} = 72616.40 Extreme**
One-sample t-test/Wilcoxon singed-rank test – Basic normalised shift, impulsive (log *k* +1)	BF_{01} = 1.12 Anecdotal	BF_{01} = 8.18 Moderate	-	-
One-sample t-test/Wilcoxon singed-rank test – Basic normalised shift, patient (log *k* -1)	**BF_{10} = 1007.63 Extreme**	**BF_{01} = 438.35 Extreme**	-	-
One-sample t-test/Wilcoxon singed-rank test – Distance-controlled normalised shift, impulsive (log *k* +1)	BF_{01} = 3.15 Moderate	BF_{01} = 8.70 Moderate	-	-
One-sample t-test/Wilcoxon singed-rank test – Distance-controlled normalised shift, patient (log *k* -1)	**BF_{10} = .19 Anecdotal support for H0**	**BF_{10} = 8.15 Moderate**	-	-

Note: BF_{01} is used for non-significant (i.e. $p > .05$) and indicates support for the null hypothesis (H0). BF_{10} is used for significant analyses ($p < .05$) and indicates support for the alternative hypothesis (H1).

Supplemental References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J. & Dolan, R. J. (2015). Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron, 85*(2), 418-428.

Vincent, B. T., & Rainforth, T. (2017). The DARC Toolbox: Automated, flexible, and efficient delayed and risky choice experiments using Bayesian adaptive design. *Preprint.*

Chapter 4.

Does Contagion Generalise Across Different Tasks?
A Comparison of Temporal and Probability Discounting

Louisa Thomas

Joshua H. Balsters

In preparation for publication

Chapter starts on **page 54** of the thesis

41 pages in chapter, including references

Abstract

Neuroeconomics paradigms have shown that learning about another's beliefs can shift the preferences of neurotypical adult participants - a process known as contagion. At present, it is unclear whether discounting (and contagion in discounting tasks) is domain specific, or domain general and shares a common framework across tasks. Here, we fit Bayesian computational models to precisely quantify choice behaviour on two tasks (i.e., temporal and probability discounting), in which participants (N = 100) make decisions on behalf of themselves and two other agents. Individuals who were more patient (temporal discounting) were also more risk-seeking (probability discounting), although participants' choices were noisier in the probability discounting task than in the temporal discounting task. Contagion effects were clearly present after observing both more patient and more risk-seeking agents. A significant correlation was found between the strength of contagion for these two agents suggesting a commonality of contagion across discounting domains. However, as in the previous study, we did not find any relationship between contagion and autistic traits.

Introduction

Bayesian associative processes have been found to underlie both learning about environmental features, as well as social values (Behrens et al., 2008; Bennett, 2015; De Martino et al., 2017; Perreault et al., 2012; Suzuki et al., 2012). Recently, Bayesian belief updating has been employed as a framework to investigate a social psychology phenomenon referred to as contagion - a form of implicit social influence whereby learning about someone else makes you more like them (Asch, 1956; Garvert et al., 2015; Thomas et al., 2021). Indeed, the integration of social information with existing choices is well captured by Bayesian models of social influence, with model comparisons indicating that Bayesian models outperform other models of social influence (De Martino et al., 2017; Moutoussis et al., 2016; Park et al., 2017). Although learning about others is crucial for development, contagion of value preferences can have a number of negative consequences. For example, financial traders may observe increased risk-seeking behaviour in their colleagues that will increase their own risk-seeking behaviour, which has the potential to contribute to financial bubbles and collapse (Suzuki et al, 2016). As such, those that are influenced by others to a lesser degree may be well-suited employees for situations where contagion of preferences could lead to negative consequences.

Within a Bayesian framework, contagion can be interpreted as the integration of an individual's prior beliefs with their beliefs about newly presented social information to form a new posterior belief. Increased autistic traits in neurotypical (NT) populations have been associated with less precise priors (Karvelis et al., 2018), and overlearning about environmental volatility has been found in autistic participants relative to NT controls (Lawson et al., 2017). Both differences in priors and estimations about the volatility of newly presented social information may contribute to differences in

contagion. Here, we measure the level of autistic traits in an NT sample to determine whether trait differences can be associated with differences in contagion. In addition, we explore contagion of value preferences in temporal (TD) and probability discounting (PD) tasks, and compare the extent of contagion across tasks.

Multiple studies have shown contagion of subjective value preference in TD and PD tasks in both NT and autistic samples (Garvert et al., 2015; Nicolle et al., 2012; Suzuki et al., 2016; Thomas et al., 2021). These tasks assess decrease in the subjective evaluation of monetary rewards as a function of time (TD), or risk (PD). Tasks assessing TD preferences involve a series of choices between small immediate rewards and larger delayed rewards, whereas tasks assessing PD involve a series of choices between small guaranteed monetary rewards and larger risky rewards (Basile & Toplak, 2015; Garvert et al., 2015; Nicolle et al., 2012; O'Connell et al., 2017; Suzuki et al., 2016). Participants in these studies assessing contagion of discounting preferences made choices on behalf of themselves, and agents with alternative value preferences. After learning the preferences of these agents, participants' own rate of discounting changed to become more like that of the other agent (Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Suzuki et al., 2016; Reiter et al., 2019).

The degree to which individuals discount future or probabilistic rewards for themselves varies in TD and PD tasks and this variability is quantified using discounting parameters (k values for TD, and α values for PD). Both TD and PD data have been found to be explained by a hyperbolic model, as the rate at which rewards are subjectively devalued decreases as the time delay increases (Ainslie, 1975; Blackburn & El-Deredy, 2013; Green & Myerson, 2004; O'Connell et al., 2017; Ohmura et al., 2006), although

exponential and linear models have also been used to explain PD data (Suzuki et al., 2016).

At present, it is unclear whether discounting is domain specific, or domain general and shares a common framework across tasks. Hayden and Platt (2007) found a significant correlation between *k* and α values, with increased preference for delayed options on TD tasks associated with an increased preference for economic risk in PD tasks in rhesus macaques, although an earlier study (Ohmura et al., 2006) found no such relationship. Meta-analyses have suggested that both the orbitofrontal cortex and ventral striatum track stimulus value independent of task, whereas regions in the parietal lobe and dorsomedial PFC appear to be selectively activated by TD and PD respectively (Bartra et al., 2013; Levy & Glimcher, 2012; Peters & Büchel, 2009). Individual differences in *k* value have been linked to variability in the ventral striatum (Hariri et al., 2006), therefore it is possible that correlations between TD and PD parameters reflect a shared valuation mechanism originating in the ventral striatum. It is possible that differences in comparing stimulus values (rather than value attribution) recruit additional frontal and parietal brain regions that are task specific (e.g., Peters & Büchel, 2009).

Whilst subjective value for the self may have a shared mechanism across tasks, recent work from Piva et al. (2019) has shown that participants learn the value preferences of others in a task specific manner. In this study, participants' behaviour on a PD task was best explained by a computational model that utilised a shared α value for self and other, whilst TD behaviour was best modelled using separate *k* values for self and other. This study aims to determine whether there are task related differences and similarities in discounting parameters, as well as in the extent of contagion across tasks.

Finding differences in contagion between tasks would support task-related differences in learning about, and being influenced by, the preferences of others.

Participants in this study completed a measure assessing autistic traits, as well as TD and PD contagion tasks, making choices on behalf of themselves, and on behalf of two modelled agents with alternative value preferences. We used a novel Bayesian method of indexing contagion to explore whether contagion occurred in both TD and PD tasks, and whether the magnitude of contagion was comparable across tasks. We also explored whether contagion was associated with autistic traits on both tasks.

We predicted that we would observe contagion effects in both samples, such that participants would shift their preferences to become more similar to another agent after learning the preferences of that agent. We also predicted that these contagion effects would be similar in magnitude across the two discounting tasks, and that contagion would be associated with autistic traits on both tasks.

Methods

Participants

One hundred NT participants (60 female, mean age = 21.35, SD $\pm$ 2.11) volunteered to the study, and were recruited from the Royal Holloway, University of London campus. This sample were also included in a collaboration study (Thomas et al., 2021). As an incentive for participation, participants were invited to enter into a prize draw to win an Amazon voucher for the amount of one of their chosen monetary outcomes (£1-20), which was drawn when data collection commenced. Informed consent was obtained, and Ethical approval was granted by the Royal Holloway Departmental Ethics Committee.

Procedure

Before the full data collection commenced, a pilot version of the PD task was carried out to determine which model best fitted this data, and how many trials would be required to achieve a stable estimate of participants discounting preferences. In the main experiment, participants completed TD and PD contagion tasks, which were presented in MATLAB using the Cogent 2000 v125 graphics toolbox, and were developed using scripts programmed by Mona Garvert (Garvert et al., 2015). In each task, participants made choices for themselves, and on behalf of two other agents. The order of the two discounting contagion tasks was counterbalanced across participants. Task order and example experimental screens for the two discounting tasks are shown in **Figure 1**. Participants also completed a block of six individual differences questionnaire measures in between the two discounting tasks, and two final questions at the end of the experiment. The total experimental protocol took around 45 minutes to complete.

Pilot task

A short pilot task was run before the main experiment, in order to determine both which model best fitted PD data, and how many trials would be required to achieve a stable estimate of discounting preferences. Participants (N = 20) completed a shortened 30-trial version of the PD task (reduced from 50 trials in Garvert et al., 2015), making choices only for themselves. Four models were fitted to the data: hyperbolic, log hyperbolic (as used in previous TD literature, e.g., Garvert et al., 2015; Thomas et al., 2021), exponential (as in Suzuki et al., 2016), and linear (also in Suzuki et al., 2016). AIC and BIC model comparisons showed that the log hyperbolic model provided the best fit to the data and was therefore used to model the choices of the others for both the

TD task (as per Garvert et al., 2015), and the PD task. Examining log α values revealed that this parameter was stable after around seven trials. The number of trials was therefore kept at 30 per block for both tasks.

Discounting tasks

Both discounting tasks were divided into five blocks of 30 trials (i.e., Self1, Other1, Self2, Other2, and Self3), with a self-timed break after 15 trials. On each trial, participants made a self-paced hypothetical choice between two monetary values (£1-20), presented simultaneously. For the TD task, the choices were between a small amount of money available immediately (e.g., £3 now), or a larger amount available after a specified delay period (i.e., tomorrow, 1 week, 2 weeks, 4 weeks, 6 weeks, 2 months, or 3 months). During the PD task, participants decided between accepting a small guaranteed amount of money (e.g., 100% chance of £3) or taking a gamble to receive a larger amount of money. The gamble option specified a percentage (i.e., 5%, 10%, 25%, 50%, 75%, 80%, 90%) indicating how likely the participant would be to receive that amount if they selected the gamble (e.g., 25% chance of £12). Responses were made using the left and right arrow keys, with the key corresponding to the location of the choice on the screen. On each trial, the side (i.e., left or right) of the guaranteed (TD) and gamble (PD) options was randomised.

Figure 1. Experimental design. A) Shows the order and type of the five blocks for each task. The Self, Other1 and Other2 colours were randomised across participants (RGB colours). B) Shows an example experimental screen for Self-trials on the TD task, participants see the type of trial (Self or name of Other), indicated by the name in colour, along with the options. The yellow box appears around the participant's selected choice, following key press. C) Shows an example experimental trial for Other-trials on the TD task, which follow the same format as Self-trials, with additional feedback. This feedback is presented on screen 500 ms after the yellow box appears, and is displayed for 1000 ms before the screen clears for the next trial. D) Shows an example experimental screen for a Self-trial in the PD task, which follows the same format as the TD task. E) Shows an example experimental screen for an Other-trial in the PD task, which also follows the same format as the TD task.

During Self-trials (blocks one, three, and five, i.e., Self1, Self2, and Self3), participants were instructed to choose for themselves according to their own preferences, as one of their choices could be selected as their potential bonus payment. Participants made choices on behalf of two simulated agents in blocks two and four (i.e., Other1 and Other2), and were informed that these were the choices of two previous participants. In fact, two gender-matched names were chosen at random (with a different allocation for each task) from a list of popular UK names to represent these two agents, and the

behaviour of these two agents was modelled within the experiment to be ±1 of the participant's own discounting parameter, with the order of the positive and negative agents counterbalanced across participants. In all Other-trials, feedback was displayed on each trial after participants had made their choice. Feedback remained on the screen for 1000 ms, and participants were informed that their choice was correct if they chose the option that would be preferred by the modelled Other, or incorrect if they chose the other option (see **Figure 1**).

Estimation of discount rates and simulation of the other agents' choices

The choices of the other agents in both tasks were modelled using a log hyperbolic model, and participants' own discounting preferences were derived using the same model. The subjective value of each choice on the TD task was calculated on each trial according to the following equation:

$$(1) \qquad V = \frac{M}{1+kD}$$

Subjective value is represented by *V*, *M* is the reward magnitude, *k* is the agent's discount rate, and *D* indicates the delay period in days. Log *k* parameter values were set between -4 and 0. Individuals with log *k* values closer to -4 are less sensitive to delay, and base their decisions more on reward magnitude (i.e., are more patient), and individuals with log *k* values closer to 0 are more sensitive to delay, discounting rewards more steeply as a function of time (i.e., are more impulsive; Carlisi et al., 2017; Garvert et al., 2015).

The subjective value of each choice on the PD task was calculated on each trial according to the following equation:

(2) $$V = \frac{M}{1+\alpha(1-P/P)}$$

Subjective value, and the reward magnitude are represented by *V*, and *M* respectively. Here, α refers to the agent's discounting parameter, and *P* refers to the probability of receiving a reward (between 0-1; i.e., 0-100%). Log α parameter values were set between -2 and 2. A log α value of 0 indicates that the individual is risk neutral and will accept rewards based solely on reward value. Values closer to +2 indicate aversion to risk (i.e., accepting more guaranteed amounts), and values closer to -2 indicate that the individual is more risk-seeking (i.e., accepts more gambles).

The value of the immediate (TD) and guaranteed (PD) options are henceforth be referred to as V_{SS}, and the value of the delayed and gamble options are referred to as V_{LL}. The subjective value of V_{SS} always corresponds to the magnitude of the reward (*M*), because the delay period in days is 0 (represented by *D* in equation (1)), or the probability attached to the reward is 0 (represented by *P* in equation (2)).

The following softmax function was used to transform the subjective values of each option (V_{SS} and V_{LL}) into choice probabilities (i.e., P_{SS} and P_{LL}).:

(3) $$P_{LL} = \frac{1}{1+e^{-\beta(V_{LL}-V_{SS})}}$$

Equation (3) shows the calculation of choice probability for P_{LL}. In this equation, β is a free parameter which characterises noise in the agent's choices. Log β parameters were set between -1 and 1. Values closer to -1 indicate greater noise in the agent's choices, with larger non-systematic deviations around the indifference point (i.e., the point at which both choice options are equally preferred).

During Other-trials, responses on each trial were modelled to reflect the preferences of an agent with a discount rate (log *k* or log α) that was ±1 of the

participant's own discounting parameter at the end of the Self1-block, calculated within the experimental script. The subjective value of each option was calculated using equation (1) or (2), and differences in subjective value of the two options were translated into choice probabilities using the softmax function shown in equation (3), with log β fixed at 1.

Participants' own log *k*, log α, and log β values were also derived from equations (1), (2), and (3), and were updated on a trial-by-trial basis using a Bayesian update rule. A uniform prior was updated on each trial to form the posterior, which was calculated as the likelihood of an individual's choice given the discounting parameter (log *k* or log α), and β, weighted by the prior. The posterior was then set as the prior for the next trial. Previous studies have shown that we use our own priors as the starting point for learning about others (Lockwood et al., 2018; Tarantola et al., 2017), hence the posterior from the end of the Self1 and Self2 blocks is used here as the prior for the start of the Other1 and Other2 blocks. To avoid the influence of strong priors for Other1 or Other2 on the estimation of priors for Self2 or Self3, the prior was reset to start with a new uniform prior at the beginning of each Self-block. The β parameter was fixed at 1 for Other-blocks, and was set to start at 0.3 for Self-blocks.

Generation of choice pairs

To ensure accurate estimations of participants' discount rates, two methods were used to generate choice pairs in all Self-blocks, and these methods were alternated across trials (as per Garvert et al., 2015 and Thomas et al., 2021). The first of these methods was a generative method. For the TD task, all possible pairs of reward amounts and delays were generated, and for the PD task, all possible pairs of amounts and percentages

were generated. For both tasks, the amount associated with the immediate or guaranteed options was always smaller than the amount associated with the delayed or gamble option. Out of all generated choice pairs, 15 trials were selected for each Self-block that best matched the indifference points of 15 hypothetical agents with log k values evenly distributed between -4 and 0, or log α values evenly distributed between -2 and 2. Choice pairs were also produced using the same method for all 30 trials in both Other-blocks, such that trials in these blocks best matched the indifference points of 30 hypothetical subjects with log k values evenly distributed between -4 and 0, or log α values evenly distributed between -2 and 2.

The remaining 15 trials in each Self-block were generated using an adaptive staircasing method, which relies on a Bayesian framework to produce precise estimations of discounting parameters. The prior was set to be normally distributed, and Bayes rule was used after each decision to update this prior. Here, the prior represents participants initial belief about discount rate, and this refines as the estimation of discount rate becomes more precise. For the TD task, this prior was initially set to be normally distributed with a mean of -2 and a standard deviation of 1, and for the PD task this was set to be normally distributed with a mean of 0 and standard deviation of 1. Choice pairs were generated which probed participants' indifference points, in order to refine the prior.

Contagion when accounting for uncertainty

A Bayesian belief update measure (Kullback-Leibler divergence (D_{KL})) was used to calculate the change in participants' beliefs about log k and log α values, after learning about the preferences of other agents. This measure quantifies the divergence in the

distribution of two data sets (Kullback & Leibler, 1951), and accounts for both the change in the peak between the two distributions, and the precision of the distributions. The change in peak reflects a change in participant's behaviour (or choice profile, expressed by log k or log α in this instance), and precision reflects the participant's certainty (Moutoussis et al., 2016). In this paper, D_{KL} quantifies the divergence in the distribution of participants' prior beliefs about log k or log α between the end trials of two blocks (i.e., the posterior). The end trials are used for this calculation, as discount rates are stable by this point, and the posterior has been updated and refined at each preceding trial.

D_{KL} values were normalised for each analysis, so that D_{KL} was positive when participant's log k or log α changed in the same direction as the behaviour of the modelled Other, or negative when the participant's log k or log α changed in the opposite direction to the Other. For example, D_{KL} would be positive if the Other's log k was +1 the log k of the participant, and the participants discount rate also increased. In this example D_{KL} would be negative if the participant's discount rate had decreased following exposure to the positive Other. Normalised D_{KL} variables were also sorted into positive (i.e., impulsive/risk-aversive) and negative (i.e., patient/risk-seeking) variables based on the direction of the behaviour of the Other (i.e., ± 1 of the participant's log k or log α).

Behavioural shift in discount rate

The shift in participant's discount rate (i.e., contagion) was also calculated according to the distance-controlled method employed by Garvert et al., (2015) to allow for direct comparison across studies, as follows:

(4) $$Shift1 = \frac{DR_{Self2} - DR_{Self1}}{DR_{Other1} - DR_{Self1}}$$

(5) $Shift2 = \frac{DR_{Self3} - DR_{Self2}}{DR_{Other2} - DR_{Self2}}$

DR is used to denote the discounting parameter of the agent. The log *k* or log α value from the last trial of each block was used. Equation (4) shows the calculation of shift at Self2 away from Self1 and towards Other1. The distance between the two Self-blocks is controlled by the distance between the participant's own discounting parameter at the end of the Self1-block, and the participant's estimate of the discount rate of the other agent, at the end of the Other1-block. Equation (5) shows this shift at Self3, away from Self2 and towards Other2. Here, the distance between the two Self-blocks is controlled by distance between the participant's own discounting parameter at the end of the Self2-block, and the participant's estimate of the discount rate of the other agent, at the end of the Other2-block.

Shift variables were also normalised (i.e., shifts in the same direction as the Other are positive, and shifts in the opposite direction are negative), and sorted into positive (i.e., impulsive/risk-aversive) and negative (i.e., patient/risk-seeking) based on the direction of the model.

Discounting lottery for bonus participant payment

From each discounting task, one trial was selected at random from all Self-trials (N = 90). The trial, and the participant's chosen outcome were shown on screen at the end of the task. If they wished, participants were then entered into a lottery at the end of the experiment to win an Amazon voucher for one of these selected amounts (i.e., one of their selected options on either a TD or a PD trial). Immediate or guaranteed amounts were sent immediately after the draw. If the participant selected a delayed amount, an email notification of the prize was sent immediately, and the voucher was sent following

the specified delay period. If the participant had selected a gamble option on the selected PD trial, a random decimal between 0 and 1 was generated. The participant won the award if the decimal equivalent of the percentage attached to their selected reward was lower than the randomly selected decimal. Following this, participants were notified whether or not they had won the reward, and they were sent the voucher if the gamble was successful.

Individual differences in social cognition self-report measures

Participants also completed a block of six individual differences questionnaire measures, which were collected for analysis as part of a collaboration project. Questionnaires were presented full screen in Qualtrics, in the following order: Social Network Index (Cohen et al., 1997), Apathy Motivation Index (Ang et al., 2017), Questionnaire of Cognitive and Affective Empathy (Reniers et al., 2011), Twenty-Item Toronto Alexithymia Scale (Bagby et al., 1994), Short Form Self-Report Psychopathy scale (Gordts et al., 2017), and the autism quotient (AQ; Baron-Cohen et al., 2001). Only AQ data are analysed in this paper, and a Likert scoring method (scored 1-4, Austin, 2005) is used for scoring the AQ, as higher internal consistency and greater test-retest reliability has been found for this scoring method in NT samples compared to the original binary scoring method (Stevenson & Hart, 2017).

Awareness of manipulation and behavioural change

To determine whether participants believed that the modelled Others were real participants, and whether they had noticed their own behaviour change during the two discounting tasks, two questions were asked following completion of all tasks and

questionnaires: 1) *"Did you believe that the other two players were real participants?"*, and 2) *"Did you notice yourself changing your choices throughout the experiment?"*.

Participant exclusions

Participant exclusions were conducted separately for the TD and PD tasks. Five outliers (i.e., ± 3 SDs of mean distance-controlled shift or mean D_{KL}) were excluded from the TD data, and four from the PD data. In the TD task, two participants with log *k* values close to zero had two modelled agents with behaviour in the same direction, as such, a more positive other could not be generated. In the PD task, there were 17 participants with the behaviour of both modelled agents in the same direction. For these participants, data were analysed for the other agent with the greatest distance between the participant's own discount rate, and the model discount rate. All 100 participants had data for either the TD or PD task included in the final analyses. In the final sample, a total of 73 participants had data included for both positive (i.e., impulsive/risk-aversive) and negative (i.e., patient/risk-seeking) agents on both tasks.

Statistical analyses

All analyses were run using Jamovi (version 1.6.15), and figures were produced in R Studio (version 1.1.463) using R (version 3.5.2). Bayes factors (calculated using Jamovi jsq package presets) quantify the evidence in favour of one hypothesis versus another (e.g., the hypothesis of interests versus the null) given the data (Kass & Raftery, 1995; Quintana & Williams, 2018). Here, Bayes factors are reported for all analyses, interpreted in line with the jsq package (version 1.0.2) criteria. BF_{10} is used for significant analyses (measured by frequentist *p* values) and denotes the likelihood that

the hypothesis of interest is correct. BF_{01} is used for non-significant analyses and denotes the likelihood that the null hypothesis is correct.

Description of all analyses

To assess the relationships between TD and PD task parameters, a correlation analyses was run between log *k* and log α values, and a paired-samples t-test was run on log β values. Exploratory analyses were also run to determine whether there was an effect of gender on task parameters (i.e., log *k* and log α values, and log β values for both tasks).

To determine whether there was a significant contagion effect in the TD and PD tasks, one-sample t-tests or Wilcoxon signed-ranks tests (depending on the distribution of each variable) were run on both normalised D_{KL} and normalised shift variables for each task, split for positive (i.e., impulsive/risk-aversive) and negative (i.e., patient/risk-seeking) others. To determine whether the shift or D_{KL} measures were more likely to produce a significant contagion effect, Bayes factors in support of contagion effects were compared for each task and direction. To assess the relationship between contagion across tasks, correlation analyses were also run between D_{KL} variables for each task and direction. To determine whether there were significant differences in contagion dependent on task or direction, a repeated-measures Task (TD/PD) x Direction (positive/negative) ANOVA was run on the D_{KL} data.

To determine whether there was a significant task or direction effect on how accurate participants were at making choices on behalf of the other (i.e., the number of trials on which the participant chose the same option as the modelled other, out of 30), a repeated-measures Task (TD/PD) x Direction (positive/negative) ANOVA was run on accuracy data. Correlation analyses were also run between normalised D_{KL} and accuracy

data to determine whether there was a significant relationship between contagion and accuracy.

In order to determine whether contagion was affected by participants' belief in the manipulation (i.e., whether or not they believed that the other agents were previous participants), or their awareness of their behaviour changing throughout the task, robust independent-measures t-tests were run, with mean normalised D_{KL} as the dependent variable, and participants answers to the final two questions as grouping variables.

Finally, two regression analyses were run to assess the association between autistic traits (as measured by the AQ) and contagion in TD and PD tasks, with data averaged across direction.

Results

Table 1. Descriptive statistics (mean (SD), and participant N for all analyses for the final sample for the TD and PD tasks.

Variable	TD			PD		
	M	*SD*	*N*	*M*	*SD*	*N*
Discounting parameter (log k/α)	-1.66	.67	95	.16	1.04	96
Log β	-.19	.35	95	-.86	.24	96
Normalised D_{KL}						
Impulsive/risk-aversive (log k/α +1)	.37	2.37	95	.17	1.22	86
Patient/risk-seeking (log k/α -1)	.94	2.45	92	.35	1.00	89
Shift						
Impulsive/risk-aversive (log k/α +1)	.80	4.39	95	.52	4.71	86
Patient/risk-seeking (log k/α -1)	.95	2.92	92	.70	2.80	89
Accuracy						
Impulsive/risk-aversive (log k/α +1)	22.36	3.39	95	22.83	2.87	86
Patient/risk-seeking (log k/α -1)	23.57	3.14	92	24.28	2.14	89

Relationship between temporal and probability discounting parameters

A significant positive correlation was found between participants' log *k* and log α values (r = .25, p = .016, BF_{10} = 2.26), indicating that individuals that were more impulsive were also more risk-averse, although Bayesian support for this finding is anecdotal.

As log β values were not normally distributed (determined by Shapiro-Wilk tests of normality, W = .94, p < .001), a Wilcoxon paired-sample test was run on participants' log β parameter values to determine whether there was an effect of task on this parameter.

This analysis revealed a significant effect of task (W (90) = 4076.00, $p < .001$, $d = 1.60$, BF_{10} = 3.010e+23). The amount of noise in participants' choices was greater in the PD task than in the TD task. See **Table 1** for descriptive statistics.

Contagion when accounting for uncertainty

To determine whether there was a significant change in behaviour after learning the value preferences of another person (i.e., contagion), as measured by normalised D_{KL} one-sample t-tests were run. If data were not normally distributed (as determined by Shapiro-Wilk tests of normality), Wilcoxon signed-rank tests were used.

For the TD task, normalised D_{KL} was significantly greater than zero for more patient agents (t (91) = 3.68, $p < .001$, $d = .38$, BF_{10} = 53.71), suggesting a significant effect of contagion after observing patient agents. Normalised D_{KL} was not significantly greater than zero for more impulsive agents (W (94) = 2609.00, $p = .223$, $d = .16$, BF_{01} = 2.86; W = .96, $p = .007$) in the TD task, although Bayesian support for the null effect was anecdotal. These findings indicate that a significant contagion effect was found for more patient (i.e., negative) agents, but not for more impulsive (i.e., positive) agents.

For the PD task, normalised D_{KL} significantly greater than zero for more risk-seeking agents (W (88) = 2850.00, $p < .001$, $d = .35$, BF_{10} = 18.82; W = .82, $p < .001$). Normalised D_{KL} was not significantly greater than zero for more risk-aversive agents in this task (W (85) = 2275.00, $p = .082$, $d = .14$, BF_{01} = 3.61; W = .88, $p < .001$). These findings indicate that a significant contagion effect was found for more risk-seeking (i.e., negative) agents, but not for more risk-aversive (i.e., positive) agents. Therefore, for both the TD and PD tasks, contagion (as measured by normalised D_{KL}) was significantly

greater than zero when observing agents with a more negative value preference (i.e. log k/α -1), but not a more positive value preference.

To determine whether contagion was related across tasks, correlation analyses were run between the normalised D_{KL} measures for each task (with critical p value adjustments made to allow for multiple comparisons in the event of a significant result; see **Table 2** for results of these analyses). Only the correlation between more negative (i.e., patient and risk-seeking) others between tasks was significant, with strong Bayesian support for the finding (see **Figure 2**).

Table 2. Correlations between D_{KL} variables for positive and negative others for both tasks.

	TD					
	Impulsive (+)			Patient (-)		
	r	*p*	*BF*	*r*	*p*	*BF*
PD						
Risk-aversive (+)	.04	.700	6.74	-.12	.288	4.11
Risk-seeking (-)	-.03	.807	7.12	**.35**	**.001***	**25.43**

Note: *BF* is BF_{10} for significant analyses and BF_{01} for non-significant analyses. * is $p <$ the adjusted p value of .013 (i.e., .050 / 4).

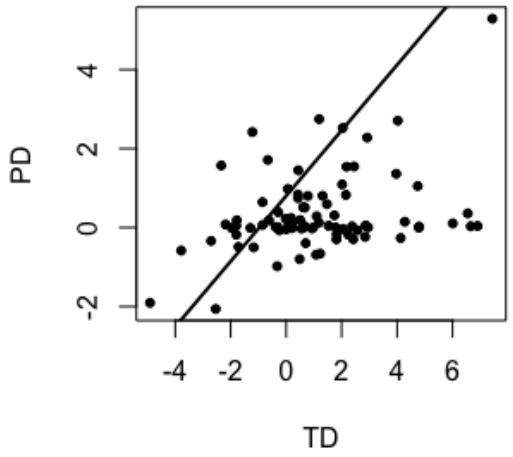


Figure 2. Correlation between contagion across tasks. This shows the correlation between contagion (as measured by normalised D_{KL}) for the negative (i.e., patient and risk-seeking) others on both discounting tasks. Positive values indicate a shift towards the preferences of the other, and negative values indicate a shift away.

To determine whether there was a significant effect of direction of the modelled agent on normalised D_{KL}, and to examine task effects, a Task (TD/PD) x Direction of the other (more positive/more negative) repeated-measures ANOVA was run on this data. The main effect of task was significant (F (1, 72) = 5.88, p = .018, η^2 = .01, BF_{10}= 1.20), with stronger contagion for the preferences of the modelled agent in the TD task than in the PD task. There was no significant main effect of direction (F (1, 72) = 1.78, p = .187, η^2 = .01, BF_{01}= 2.38), and no significant interaction effect (F (1, 72) = .90, p = .347, η^2 = .00, BF_{01}= 1.82). However, Bayesian support for all results of this ANOVA are anecdotal. See **Figure 3** for all normalised D_{KL} variables.

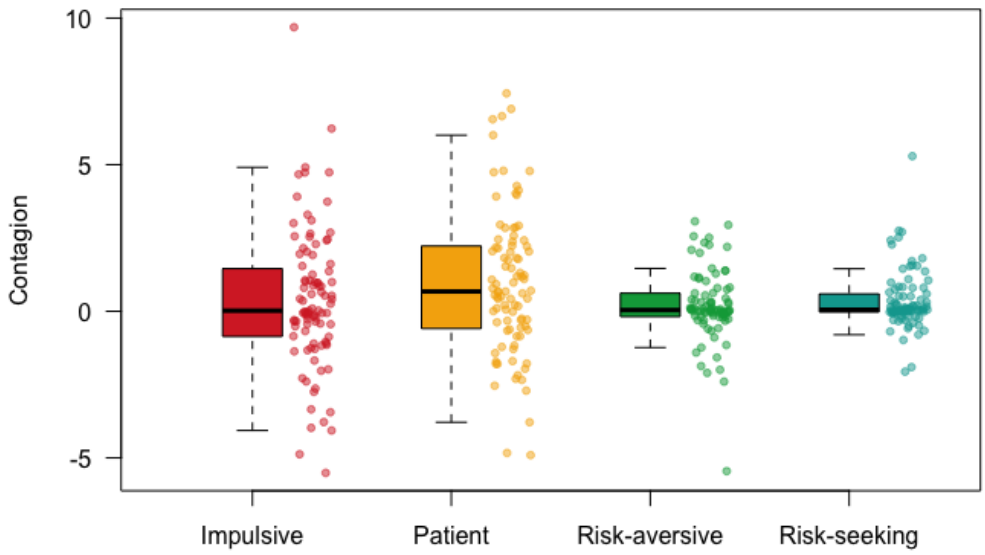


Figure 3. Task and direction comparisons of normalised D_{KL} scores: Descriptive statistics for all divergence score (D_{KL}) variables for the two discounting tasks, split for direction (i.e., positive, impulsive/risk-aversive and negative, patient/risk-seeking). Positive values indicate a change towards preferences of the other agent (i.e., contagion), whereas negative values indicate a shift away.

Normalised behavioural shift in discount rate towards the modelled Other

To determine whether there was a significant shift in participant's preferences between Self-blocks (in order to replicate the findings of previous research by Garvert et al., 2015), one-sample t-tests were run on normalised shift variables. If data were not normally distributed (as determined by Shapiro-Wilk tests of normality), Wilcoxon signed-rank tests were used.

For the TD task, normalised shift was significantly greater than zero for the more patient agent (W (91) = 3129.00, $p < .001$, $d = .33$, $BF_{10} = 10.64$; W = .80, $p < .001$). Normalised shift was not significantly greater than zero for the more impulsive agent (W (94) = 2573.00, $p = .278$, $d = .18$, $BF_{01} = 1.95$; W = .78, $p < .001$), with anecdotal Bayesian support for the null effect.

For the PD task, normalised shift was significantly greater than zero for more risk-seeking agents (W (88) = 2893.00, $p < .001$, $d = .25$, $BF_{10} = 1.55$; W = .72, $p < .001$), with anecdotal Bayesian support for the effect. Normalised shift was not significantly greater than zero for the more risk-aversive agent (W (85) = 2168.50, $p = .200$, $d = .11$, $BF_{01} = 5.08$; W = .77, $p < .001$). These findings indicate that a significant contagion effect was observed for more negative (i.e., patient/risk-seeking) agents, but not more positive (i.e., impulsive/risk-aversive) agents, when contagion is measured by normalised shift, and normalised D_{KL}.

Comparison of shift and D_{KL} measures of contagion

Bayes factors in support of an effect of contagion (i.e., BF_{10}) were compared for the normalised shift and D_{KL} variables for each task and direction, in order to determine whether an effect of contagion was more likely when accounting for uncertainty (i.e.,

D_{KL}) or when using the shift measure (see **Table 3** for Bayes factors). For the TD task, an effect of contagion was 1.47 times larger for the positive/impulsive other when using the normalised shift measure versus the normalised D_{KL} measure. However, contagion was 5.05 times larger for the negative/patient other when using the normalised D_{KL} measure versus the normalised shift measure.

For the PD task, an effect of contagion was 1.41 times larger for the positive/risk-aversive other when using the normalised D_{KL} measure versus the normalised shift measure, and 12.12 times larger for the negative/risk-seeking other when using the normalised D_{KL} measure versus the normalised shift measure. These findings indicate than an effect of contagion was greater when using the D_{KL} measure and accounting for uncertainty versus the shift measure in the majority of conditions.

Table 3. Bayes factors (BF_{10}) in support of an effect of contagion for normalised shift and D_{KL} variables for each discounting task and direction.

	Variable	
	Normalised shift	Normalised D_{KL}
TD		
Impulsive (+)	.513	.349
Patient (-)	10.644	53.710
PD		
Risk-aversive (+)	.197	.277
Risk-seeking (-)	1.553	18.824

Accuracy when making choices for the Other

To examine the effects of task and direction of the behaviour of the Other on accuracy, a Task (TD/PD) x Direction (more positive/more negative) repeated-measures ANOVA was run on this data (see **Figure 4**). The main effect of task on accuracy was

significant (F (1, 72) = 5.39, p = .023, η^2 = .01, BF_{10} = .88). Participants were more accurate at making choices on behalf of the other agent in the PD task than in the TD task, although Bayesian support for this finding was anecdotal. There was also a significant main effect of direction (F (1, 72) = 19.42, p < .001, η^2 = .07, BF_{10} = 7277.63), such that participants were more accurate at making choices on behalf of agents with increasingly negative discounting parameters (i.e., more patient/risk-seeking others) than agents with more positive discounting parameters (i.e., more impulsive/risk-aversive others). The interaction between task and direction was not significant (F (1, 72) = .02, p = .889, η^2 = .00, BF_{01} = 1.326e-4).

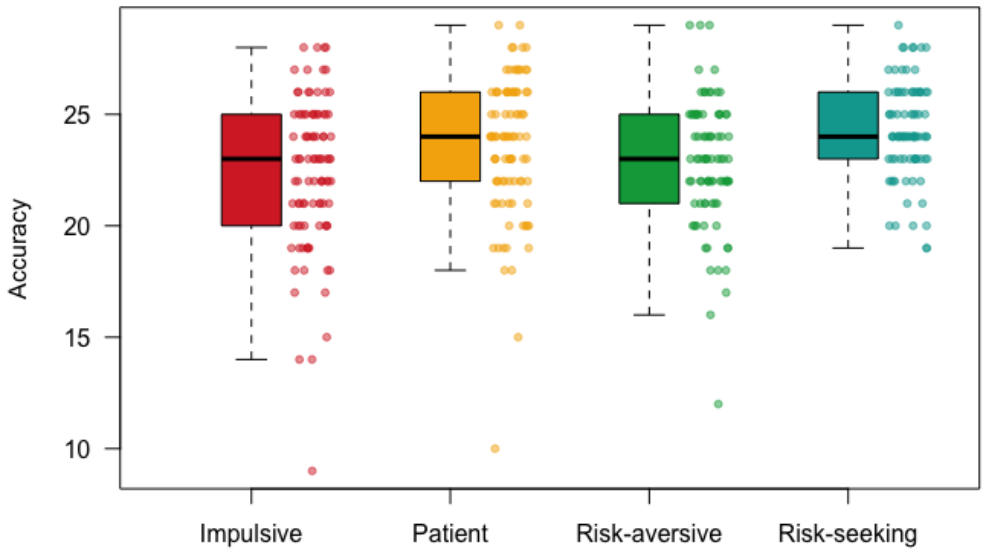


Figure 4. Task and direction comparisons of accuracy: Descriptive statistics for all accuracy variables (accuracy scored out of 30) for the TD and PD tasks, split for direction (i.e., positive, impulsive/risk-aversive and negative, patient/risk-seeking).

To determine whether contagion was related to accuracy, correlation analyses were run between the normalised D_{KL} and accuracy measures for each task and direction (with critical p value adjustments made to allow for multiple comparisons in the event of a significant result; see **Table 4** for results of these analyses). Only the relationship

between accuracy and D_{KL} for the more impulsive other on the TD task was significant, with anecdotal Bayesian support. This suggests that the relationship between accuracy and contagion was not reliable.

Table 4. Correlation matrix with correlations presented between accuracy and D_{KL} for each direction on the TD and PD tasks.

		D_{KL} TD		D_{KL} PD	
		Impulsive (+)	Patient (-)	Risk-aversive (+)	Risk-seeking (-)
Accuracy TD					
Impulsive (+)	*r*	**.26**	-	-	-
	p	**.012***	-	-	-
	BF	**2.86**	-	-	-
Patient (-)	*r*	-	-.15	-	-
	p	-	.148	-	-
	BF	-	2.73	-	-
Accuracy PD					
Risk-aversive (+)	*r*	-	-	.16	-
	p	-	-	.155	-
	BF	-	-	2.75	-
Risk-seeking (-)	*r*	-	-	-	.04
	p	-	-	-	.744
	BF	-	-	-	7.16

Note: *BF* is BF_{01} for non-significant analyses, and BF_{10} for significant analyses. * denotes a significant findings at the adjusted critical *p* value of .013 (i.e., .050 / 4).

Awareness of manipulation and change, and effect on contagion

A total of 89 participants completed the awareness of manipulation question (39.33% believed manipulation), and 76 participants completed the change awareness question (56.58% noticed their behaviour changing), indicating that the majority of

participants did not believe that they were making choices on behalf of a real participant, and that the percentage of participants that noticed their behaviour was changing was close to chance. To determine whether participant's answers to these questions were related to contagion, four robust independent measures t-tests were run, on mean normalised D_{KL} variables.

For the TD data, there was no significant effect of awareness of manipulation on mean normalised D_{KL} (t (41.70) = 1.56, p = .127, ξ = .28, BF_{01} = 1.80). There was also no significant effect of awareness of behavioural change on mean normalised D_{KL} (t (24.16) = 1.51, p = .144, ξ = .25, BF_{01} = 1.39). For the PD data, there was also no significant effect of awareness of manipulation (t (35.07) = 1.05, p = .300, ξ = .19, BF_{01} = 3.19), or awareness of behavioural change (t (30.34) = 1.01, p = .321, ξ = .18, BF_{01} = 2.58) on mean normalised D_{KL}. These results indicate that participants were influenced by the choices of the other to the same degree, regardless of whether they believed in the manipulation, or noticed their behaviour changing.

Gender differences in individual task parameters

We also conducted exploratory analyses to determine whether there were gender differences in task parameters (see **Table 5** for descriptive statistics). Independent-measures t-tests revealed no significant gender difference in log k (t (93) = 1.30, p = .197, d = .27, BF_{01} = 2.17), although the gender difference in log α was significant (t (94) = 2.18, p = .032, d = .46, BF_{10} = 1.74), with more positive/risk-aversive values for males than females, although Bayesian support for this effect was anecdotal. There was a significant gender difference in log β for the TD task (t (93) = 2.44, p = .016, d = .51, BF_{10} = 2.91), with anecdotal Bayesian support, whereby females made noisier choices

than males. There was no significant difference in log β for the PD task (t (94) = .75, p = .458, d = .16, BF_{01} = 3.58).

Table 5. Descriptive statistics (mean (SD)) for task parameters for the TD and PD tasks, split by gender.

	TD		PD	
	Log k	Log β *	Log α *	Log β
Male	-1.55 (.76)	-.08 (.37)	.44 (1.02)	-.83 (.30)
Female	-1.74 (.60)	-.26 (.32)	-.03 (1.02)	-.87 (.19)

Note: * indicates significant analyses.

Relationship between AQ subscales and contagion

To determine whether there were any significant relationships between D_{KL}, and the subscales of the AQ, data were collapsed across the two directions for each task (collapsed data were used because the effect of the direction of the other on D_{KL} was not significant), and two regression analyses were run (see **Table 6** for descriptive statistics for the AQ data and regression statistics). In each regression analysis, mean normalised D_{KL} is entered as the outcome variable, and the predictors are the five subscales of the AQ: Social Skills, Attention Switching, Attention to Detail, Communication, and Imagination, scored used the Likert scoring method.

Whilst the majority of participants in our sample scored below 32 (80% of diagnosed adults scored 32+ in the initial study; Baron-Cohen et al., 2001), two participants in our sample scored 32 or over. However, this is 2% of our total sample, and 2% of controls also scored over 32 in the Baron-Cohen et al. (2001) study, indicating that our sample is in line with AQ scores found in the general population.

For the TD task, the overall model explained -.05% of the variance in mean normalised D_{KL}, and was not a significant predictor overall (F (5, 87) = .19, p = .966,

BF_{01} = 123.55). For the PD task the overall model explained .04% of the variance in mean normalised D_{KL}, and was also not a significant predictor overall (F (5, 88) = 1.68, p = .147, BF_{01} = 7.45). Only the Attention to Detail subscale significantly predicted mean normalised D_{KL} in the PD task, although the correlation between the two variables was not significant (r = -.03, p = .790, BF_{01} = 7.49). None of the remaining predictors were significant, indicating that there may not be a reliable relationship between autistic traits and contagion in NT individuals for either TD (as also demonstrated in Thomas et al., 2021) or PD tasks.

Table 6. Descriptive statistics (mean (SD)) for the binary scoring method and the Likert scoring method for the AQ subscales and total score are presented here, along with regression statistics for the individual predictors (AQ subscales) entered into the regression models to predict mean normalised D_{KL} on the TD and PD tasks.

			TD			PD		
	Binary	Likert	β	t	*p*	β	t	*p*
AQ								
Social skills	2.52 (1.99)	20.62 (3.99)	.03	.49	.623	.01	.26	.798
Attention switching	4.55 (2.14)	24.39 (3.97)	.01	.20	.846	.06	2.64	.010*
Attention to detail	5.13 (2.32)	24.93 (4.37)	-.03	-.71	.480	-.00	-.24	.812
Communication	2.21 (1.97)	19.83 (3.79)	.00	.03	.976	-.04	-1.53	.130
Imagination	2.28 (1.82)	19.81 (4.02)	-.00	-.08	.937	.01	.25	.800
TOTAL	16.69 (6.63)	109.57 (13.09)	-	-	-	-	-	-

Discussion

Here, we used a novel Bayesian method of indexing contagion to explore whether contagion of value preferences occurred for both TD and PD tasks, and whether the magnitude of contagion was comparable across tasks. We found contagion of value preference for both tasks, suggesting a commonality of contagion across multiple domains.

Individuals who were more patient (increasingly negative *k* value) on the TD task were also more risk-seeking (increasingly negative α value) on the PD task. This is in keeping with previous studies suggesting the existence of a common-currency framework for discounting tasks (Bartra et al., 2013; Hayden & Platt, 2007; Levy & Glimcher, 2012; Wittman et al., 2018). However, Bayesian support for this finding was anecdotal, and this finding has also been challenged by other studies that have suggested distinct mechanisms underpin TD and PD (Myerson et al., 2003; Peters & Büchel, 2009; Piva et al., 2019). Our results also support task differences as we demonstrated clear differences in participants' own log β values between tasks, with larger values for the PD task than the TD task, suggesting that PD choices were noisier and more variable than TD choices. Whilst there may be task-related relationships between discounting parameters, or similarities in how these are tracked in the brain (Bartra et al., 2013; Hayden & Platt, 2007; Levy & Glimcher, 2012; Wittman et al., 2018), our results highlight the importance of distinguishing between value assignment (i.e., discounting parameters) and choice behaviour (i.e., β values) in assessing similarities and differences in discounting behaviour across tasks.

In line with previous research, we found that learning the different discounting preferences of another agent led to significant shifts in the participant's own discount

rate (i.e., contagion; Garvert et al., 2015; Nicolle et al., 2012; Suzuki et al., 2016, **chapter 3**, Thomas et al., 2021). Contagion effects were strongest after observing both more patient and more risk-seeking agents (i.e., agents with increasingly negative discounting parameters). The bias towards becoming more patient was also demonstrated by Moutoussis and colleagues (2016), and Suzuki and colleagues (2016) also found slightly larger contagion effects for risk-seeking others versus risk-aversive others (Suzuki et al., 2016). It is possible that in both cases, contagion was greater in one direction because there was greater room for change. For PD, the majority of adults tend to be a little risk-aversive rather than risk neutral (Paulsen et al., 2012), and thus behaviour which is more risk-seeking is more noticeably different from participant's own behaviour. In the TD task, we set boundaries of log *k* between 0 and -4, as in previous research (Garvert et al., 2015; Moutoussis et al., 2016). However, two participants had log *k* values so close to zero that a more positive other could not be generated, the mean log *k* value in our sample was also -1.66, which indicates that most participants were closer to impulsive than patient, also leaving more room for change when observing a patient agent. However, research exploring social influence in a PD paradigm showed that participants who were more risk-aversive, were more influenced by risk-aversive others, and participants who were more risk-seeking, were more influenced by risk-seeking others (Chung et al., 2015). Future research could aim to explore whether varying the difference between the discounting parameters of the participants and other agents would impact upon contagion. Doing so would produce agents with behaviour that is less noticeably different from that of the participant, and could help to determine whether contagion is affected by the perceived similarity of the other in these tasks.

A significant correlation was also found between contagion for more negative others on the TD and PD tasks, suggesting that contagion occurs in a similar manner across tasks. Whilst the findings of Piva et al. (2019) suggest that PD behaviour was best explained by a model using shared discounting parameters for self and other, and TD behaviour was best explained by a model using separate discounting parameters for self and other, it appears that the nature of value representation for the other agent (e.g., using a shared or independent discounting parameter) is unrelated to the extent of behavioural change. As a significant correlation between contagion was only observed for conditions in which there was also a significant contagion effect, future research exploring the effects of distance between self and other on contagion should also explore whether contagion is related across tasks if a significant effect is also found for agents with more positive discounting parameters.

We also found that contagion was significant for both discounting tasks using both the shift method previously employed by Garvert et al. (2015) as well as our novel approach using D_{KL} to examine the shift in priors. Comparison of Bayes factors revealed that the magnitude of the contagion effect was smaller when using the normalised shift variable (i.e., the weighted average of the distribution) compared to normalised D_{KL} in the majority of conditions. This suggests that the alternative hypothesis (i.e., a significant change in value preference) was more likely to be detected using D_{KL} than the shift method used by Garvert and colleagues (2015). As D_{KL} accounts for both precision and overall change in behaviour, it is possible that incorporating additional information into the measure of contagion produces a stronger result and is more appropriate given that it incorporates uncertainty into the analysis of behavioural changes.

Subscale scores on the AQ were not predictive of contagion in the TD and PD tasks, indicating no relationship between autistic traits and contagion. Whilst previous research into social influence in autism has focused on conformity, with varied findings (Lazzaro et al., 2018; Van Hoorn et al., 2017; Yafai et al., 2014), the finding of no relationship between autistic traits and contagion is in line with the findings of Lazzaro et al. (2018), and Bowler and Worley (1994), who explored social influence effects in autistic and NT samples, and found no significant group differences. This supports the notion that social influence may be unrelated to autistic traits both in NT and autistic samples.

However, it is possible that the effect being examined here is not truly social, and is a general informational effect of learning about another agent, regardless of the identity of the agent (Garvert et al., 2015). Previous research has found no difference in contagion between tasks in which the other agent was believed to be human, or a computer agent (Garvert et al., 2015). In the current task, the majority of participants also did not believe that the other was a real participant, and we were unable to associate individual differences in belief that the agent was real with the size of the contagion effect. Varying the social closeness of the other agents in these tasks could aid in determining whether contagion is social, and thus affected by factors such as social distance, or whether the effect is brought about purely by learning of value.

Taken together, these findings demonstrate that learning the value preferences of other agents changes our own value preferences (i.e., contagion). There was no reliable evidence of a relationship between autistic traits and contagion on either task, and there was also no relationship between belief in the manipulation and contagion, suggesting that contagion occurred regardless of the believed identity of the other agent. These

findings also demonstrate that there are both similarities and differences in contagion and task parameters across discounting tasks, and highlights the importance of distinguishing between discounting parameters, and overall noise in individual choice behaviour.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*(4), 463-496.

Ang, Y. S., Lockwood, P., Apps, M. A. J., Muhammed, K., & Husain, M. (2017). Distinct subtypes of apathy revealed by the Apathy Motivation Index. *PLoS ONE, 12*(1), 1–15.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1–70.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences, 38*(2), 451–460.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale - Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427.

Basile, A. G., & Toplak, M. E. (2015). Four converging measures of temporal discounting and their relationships with intelligence, executive functions,

thinking dispositions, and behavioral outcomes. *Frontiers in Psychology, 6*, 728.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*(7219), 245–249.

Bennett, D. (2015). The neural mechanisms of Bayesian belief updating. *The Journal of Neuroscience, 35*(50), 16300–16302.

Blackburn, M., & El-Deredy, W. (2013). The future is risky: Discounting of delayed and uncertain outcomes. *Behavioural Processes, 94*, 9–18.

Bowler, D. M., & Worley, K. (1994). Susceptibility to social influence in adults with Asperger's syndrome: A research note. *The Journal of Child Psychology and Psychiatry, 35*(4), 689-697.

Carlisi, C. O., Norman, L., Murphy, C. M. Christakou, A., Chantiluke, K., Giampietro, V., … Rubia, K. (2017). Comparison of neural substrates of temporal discounting between youth with autism spectrum disorder and with obsessive-compulsive disorder. *Psychological Medicine, 47*(14), 2513-2527.

Chung, D., Christopoulos, G., King-Casas, B., Ball, S. B., & Chiu, P. H. (2015). Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. *Nature Neuroscience, 18*, 912–916.

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *JAMA: The Journal of the American Medical Association, 277*(24), 1940–1944.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgments according to its reliability. *The Journal of Neuroscience, 37*(25), 6066–6074.

Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J., & Dolan, R. J. (2015). Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron, 85*(2), 418–428.

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & Rossi, G. (2017). Validity of the Self-Report Psychopathy Scales (SRP-III full and short versions) in a community sample. *Assessment, 24*(3), 308–325.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*(5), 769–792.

Hariri, A. R., Brown, S. M., Williamson, D. E., Flory, J. D., de Wit, H., & Manuck, S. B. (2006). Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *The Journal of Neuroscience, 26*(51), 13213–7.

Hayden, B. Y., & Platt, M. L. (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. *Current Biology, 17*(1), 49–53.

Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriè, P. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife, 7*, 1–29.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience, 20*(9), 1293–1299.

Lazzaro, S. C., Weidinger, L., Cooper, R. A., Baron-Cohen, S., Moutsiana, C., & Sharot, T. (2018). Social conformity in autism. *Journal of Autism and Developmental Disorders, 49,* 1304-1315.

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.

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications, 9*(1), 4747.

Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of inter-temporal preferences. *PLoS Computational Biology, 12*(7), 1–17.

Myerson, J., Green, L., Scott Hanson, J., Holt, D. D., & Estle, S. J. (2003). Discounting delayed and probabilistic rewards: Processes and traits. *Journal of Economic Psychology, 24*(5), 619–635.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114–21.

O'Connell, G., Hsu, C. T., Christakou, A., & Chakrabarti, B. (2017). Thinking about others and the future: Neural correlates of perspective taking relate to preferences for delayed rewards. *Cognitive, Affective, & Behavioral Neuroscience,* 1–8.

Ohmura, Y., Takahasi, T., Kitamura, N., & Wehr, P. (2006). Three-month stability of delay and probability discounting measures. *Experimental and Clinical Psychopharmacology, 14*(3), 318-328.

Park, S. A., Goïame, S., O'Connor, D. A., & Dreher, J.-C. (2017). Integration of individual and social information for decision-making in groups of different sizes. *PLoS Biology, 15*(6), 3–28.

Paulsen, D. J., Platt, M. L., Huettel, S. A., & Brannon, E. M. (2012). From risk-seeking to risk-averse: The development of economic risk preference from childhood to adulthood. *Frontiers in Psychology, 3,* 313.

Perreault, C., Moya, C., & Boyd, R. (2012). A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior, 33*, 449–459.

Peters, J., & Büchel, C. (2009). Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. *The Journal of Neuroscience, 29*(50), 15727–34.

Piva, M., Velnoskey, K., Jia, R., Nair, A., Levy, I., & Chang, S. W. (2019). The dorsomedial prefrontal cortex computes task-invariant relative subjective value for self and other. *eLife, 13*(8), 1–29.

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry, 18*(1), 178.

Reiter, A. M. F., Suzuki, S., O'Doherty, J. P., Li, S-C., & Eppinger, B. (2019). Risk contagion by peers affects learning and decision-making in adolescents. J*ournal of Experimental Psychology: General, 148*(9), 1494–1504.

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2009). The QCAE: A questionnaire of cognitive and affective empathy. *European Psychiatry, 24*(1), 840.

Stevenson, J. L., & Hart, K. R. (2017). Psychometric properties of the autism-spectrum

quotient for assessing low and high levels of autistic traits in college students. *Journal of Autism and Developmental Disorders, 47*(6), 1838–1853.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., … Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron, 74,* 1125–1137.

Suzuki, S., Jensen, E. L. S., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioral contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *PNAS, 113*(14), 3755–60.

Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications, 8*(1), 817.

Thomas, L., Lockwood, P. L., Garvert, M. M., & Balsters, J. H. (2021). Contagion of temporal discounting value preferences in neurotypical and autistic adults. *Journal of Autism and Developmental Disorders.* Advance online publication.

Van Hoorn, J., Van Dijk, E., Crone, E. A., Stockmann, L., & Rieffe, C. (2017). Peers influence prosocial behavior in adolescent males with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 47*(7), 2225–2237.

Wittmann, M. K., Lockwood, P. L., & Rushworth, M. F. S. (2018). Neural mechanisms of social cognition in primates. *Annual Review of Neuroscience, 41*(1), 99–118.

Yafai, A.-F., Verrier, D., & Reidy, L. (2014). Social conformity and autism spectrum disorder: A child-friendly take on a classic study. *Autism, 18*(8), 1007–1013.

Chapter 5.

The Power of Friendship: Does Social Distance Modulate Neural Pathways Controlling Social Contagion and Learning?

Louisa Thomas
Joshua H. Balsters

In preparation for publication

Chapter starts on **page 55** of the thesis

43 pages in chapter, including supplemental information and references

Abstract

Previous research has shown that learning the value preferences of another individual changes our own value preferences (i.e., contagion). However, research has yet to determine whether this effect is truly social or is purely informational. Here, we fit Bayesian computational models to a temporal discounting contagion task, where 16 pairs of participants (close friends or partners) made decisions for themselves, on behalf of their testing partner, and a stranger. A significant effect of contagion was found, whereby participant's own behaviour changed to become both more like that of their close testing partner, and the stranger, although there was no difference in contagion for close versus distant others, and we found no significant activation related to contagion in any brain area. A social distance effect was found on learning and feedback processing, with greater accuracy for close versus distant others, greater activation in the pMFC related to surprise, and greater activation in the postcentral gyrus in response to feedback for close versus distant others. These findings demonstrate that learning and feedback signals on this task differ depending on the identity of the other agent, and future research could further test for a social distance effect on contagion using different experimental protocol or analysis methods.

Introduction

Our behaviour and preferences have been shown to be biased by social influence (Behrens et al., 2008; Cialdini & Goldstein, 2004; Izuma, 2013; Meyer et al., 2019; Raafat et al., 2009; Shamay-Tsoory et al., 2019). Multiple studies have shown that learning about the value preferences of others changes our own value preferences, and this is referred to as contagion (Garvert et al., 2015; Nicolle et al., 2012; Suzuki et al., 2016; Reiter et al., 2019; Thomas et al., 2021). However, little is currently known about why some individuals are more susceptible to an effect of contagion than others, and few studies have explored whether the degree of social contagion varies depending on the identity of the other agent. It's important to know whether people are more influenced by those that they are close to. For example, a recent study indicates that propensity to follow social distancing guidelines and wear masks during the COVID-19 pandemic is affected by the behaviour of others, with individuals being more likely to follow social distancing and wear masks if those that they were close to also did (Tunçgenç et al., 2021). Here, we use fMRI and Bayesian modelling to investigate the neural basis of learning the value preferences of others, and contagion of value preference in an adult sample. We also sought to determine whether learning and contagion are affected by the relationship between the participant and their testing partners (i.e., social distance).

Previous research has found changes in participants' temporal discounting (TD) value preferences after learning the preferences of others (e.g., Apps & Ramnani, 2017; Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Thomas et al., 2021). TD tasks involve multiple inter-temporal choice trials between small immediate rewards and larger delayed rewards and assess decrease in the subjective valuation of reward as a function of time (Myerson & Green, 1995). After successfully learning the value

preferences of the other individual (or other group), participants' own TD value preferences changed to become significantly more like those of the other (Apps & Ramnani, 2017; Garvert et al., 2015; Nicolle et al., 2012; Thomas et al., 2021). Previous research has indicated that contagion is driven by both the credibility of social information (Behrens et al., 2008; De Martino et al., 2017), and the participant's degree of certainty in their initial beliefs (Moutoussis et al., 2016), which is well captured by Bayesian models of social influence. Here, and in line with previous research (Moutoussis et al., 2016; Thomas et al., 2021), we used a Bayesian method to index behavioural change (i.e., contagion).

The change in TD value preferences has been found to be reflected in multiple areas, predominantly in the medial prefrontal cortex (mPFC; Garvert et al., 2015), which has been associated with reward valuation for both self (Boorman et al., 2009; Hunt et al., 2012; Kable and Glimcher, 2007), and other (Jenkins et al., 2008; Nicolle et al., 2012). This area also appears to play a key role in representing connections with others (Courtney & Meyer, 2020). Participants in the study by Courtney and Meyer (2020) completed trait judgments for self, close others, acquaintances, and celebrities, and rated their feelings of closeness for each individual. Examining multivariate activation patterns associated with each target revealed that these were clustered into three social categories: self, social network members (i.e., close others and acquaintances), and celebrities. Activation in the mPFC also increased linearly with increasing social closeness. Thus, it appears that a region implicated in contagion (mPFC) processes both rewards for self and other and distinguishes between social categories.

Despite this, it is possible that contagion in this task is a general informational effect of learning about another agent and is not necessarily social. For example, previous

research has found no difference in the shift in participants' preferences depending on whether they were told they were deciding for a human or a computer partner (Garvert et al., 2015), and whether they believed that the human partner was real or not (Thomas et al., 2021). Recent research has explored the influence of social distance on contagion of probability discounting (PD) value preferences in male adolescent and adult samples (Reiter at al., 2019). Participants made PD choices for themselves and observed and predicted the choices of same age peers and non-peers. Adolescents showed significantly stronger contagion for more risk-seeking peers than non-peers. Adults were also more influenced by risk-seeking peers than non-peers, but this difference was not significant. However, whilst participants in the Garvert et al. (2015) and Reiter et al. (2019) studies were introduced to the testing partners, they were still strangers to the participants. It is possible that more naturalistic conditions, with participants making TD choices on behalf of real-life close others, we may observe a significant effect of social distance in adult samples.

The change or update in TD value preference has been found to be driven by learning (measured by accuracy; Reiter et al., 2019), as well as the strength of prediction error (PE) signalling (Garvert et al., 2015; Suzuki et al., 2016). Garvert et al. (2015) also found that simulation of the other's choices was required for this shift in preference. Together, these findings indicate that learning the preferences of others is required in order to drive a contagion effect. Previous research has found that learning signals are affected by social distance, for example, in the ventral striatum, greater responses to feedback have been observed for questions relating to in-group versus outgroup members (Powers et al., 2016), and activity in this area was greater when rewards were shared with friends versus confederates or computers (Fareri et al., 2012). Greater PE signals have

also been observed in the ACC in response to the errors of friends versus strangers on a Stroop task (Kang et al., 2010), and larger error-related potentials in response to the errors of others have been associated with increased social similarity on a flanker task (Carp et al., 2008). As contagion is driven by learning signals, it is of interest to determine whether both learning related activity (e.g., in response to feedback) and contagion vary dependent on the identity of the other agent.

Here, we aimed to determine whether social distance affected behavioural and neurobiological measures of contagion and learning about others using real-world pairs of close others, and strangers. Finding an effect of social distance on contagion of TD preferences could indicate that there is a social element to the contagion effects observed in previous research. It is predicted that learning of the TD preferences of the other will produce shifts in behaviour (i.e., contagion) towards both agents, with stronger contagion effects, and more effective learning for close others versus strangers. It is also predicted that brain activation related to contagion, and feedback responses will be greater for close versus distant others.

Methods

Participants

A sample of 32 right-handed participants were recruited in pairs from the Royal Holloway campus. The pairs of participants were formed of close friends or romantic partners that had known each other for six months or more (mean length of relationship in years: 4.88, SD ± 8.66, range: 0.5-34, see **Figure 1**). Participants had not had any contact with confederates prior to the experiment. Two participants were excluded due

to excessive head movements (>5mm) in the fMRI task (final sample 24 female and six male, mean age: 21.60, SD ± 5.05).

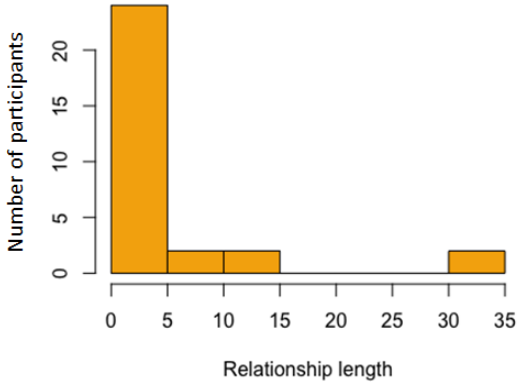


Figure 1. Participant relationship with Close Other. This shows the length of relationship with the close Other for all participants.

Participants were told that they would be paid the amount of one of their monetary choices from the fMRI task, but all participants were paid a randomly selected amount between £10 and £15 following completion of all tasks. Informed consent was obtained from participants, and additional consent was obtained for participation in an MRI study in line with the Combined Universities Brain Imaging Centre (CUBIC; http://www.pc.rhul.ac.uk/sites/cubic/) requirements. Ethical approval was obtained from the Royal Holloway Departmental Ethics Committee, and the study protocol corresponded to the CUBIC MRI Rules of Operations.

MRI apparatus

Subjects lay supine in an MRI scanner with the first and second fingers of their right hands positioned on the sixth and seventh keys of a two-handset MRI-compatible response box, with five buttons per hand. The task was projected onto a screen behind the participant and viewed in a mirror positioned above the participant's face. Event

timings and reaction times were calculated within the MATLAB experimental script, which was set to begin with a scanner pulse, and recorded button presses from the response box in the scanner.

Task order

Participants first attended a behavioural training session in their pairs alongside a stranger (confederate, i.e., Distant Other). Both participants first completed a short questionnaire collecting demographics and assessing the closeness of their relationship with their testing partner (Close Other) and the Distant Other. Participants were asked how they defined their relationship with their testing partner (i.e., *"How would you define your relationship with your close other?"*) and how long they had known each other (i.e., *"How long have you known this person?"*). Each participant then completed the one-item Inclusion of Other in the Self (IOS) scale (Aron et al., 1992). Participants select one of seven pairs of circles (labelled *"Self"* and *"Other"*) that range from just touching, to almost completely overlapping, and this item is scored from 1 (no overlap) to 7 (almost complete overlap), with 7 indicating the highest level of closeness. Participants completed one version of the IOS for their relationship with their Close Other, and one for their relationship with the Distant Other.

Both the participants and the confederate completed a short version of the TD task, making choices only for themselves. Participants were told that these choices would then be used in the scanning session, although the choices of the Close and Distant Others were modelled within the experimental script during this session. All participants completed the task in the same room, in order to increase the belief that the choices of the Others in the fMRI task were real. After the training task, the two participants entered

the scanner individually. Whilst in the scanner, participants first had a five-minute structural scan, followed by the TD contagion task, in which they made TD choices on behalf of themselves, their Close Other, and the Distant Other.

Two questions (with *"yes"* or *"no"* responses) were asked following the scanning session, to determine whether participants believed that the choices they observed in the scanner were real (*"Did you believe that the choices of the other two players were real?"*), and whether they were aware of their behaviour changing during the experiment (*"Did you notice yourself changing your choices throughout the experiment?"*).

Behavioural training

The training task consisted of 50 TD trials presented in MATLAB using PsychToolbox and was developed based on the experimental script programmed by Mona Garvert (using the Cogent 2000 v125 graphics toolbox; Garvert et al., 2015). Participants made hypothetical choices between receiving a small amount of money (£1-19) immediately and a larger amount (£2-20) available after a specified delay (i.e., tomorrow, one week, two weeks, four weeks, six weeks, two months, or three months).

Participants saw three screens in each trial (see **Figure 2**). The first screen showed the two options, one on top of the other, displayed for two seconds (Offer Stage). At this stage, participants were not required to respond, and only needed to think about which item they would prefer. Participants then viewed the same options presented side by side and were required to make a response within two seconds (Choice Stage). Responses were made on a computer keyboard, using the first and second fingers of the right hand (to mimic the actions that would be required to make a response in the scanning session),

with each finger corresponding to the location of the choice on the screen (i.e., first finger for left, and second finger for right). The location (i.e., top or bottom and left or right) of the immediate option was randomised on a trial-by-trial basis. The final screen provided feedback to participants (Feedback Stage). If the participant had made their selection in time, the text of the selected option changed to yellow for one second. If the participant missed the response, the phrase *"Response missed!"* appeared on the screen in red for one second. Following feedback, the next trial began. Between the Offer Stage and Choice Stage screens, and following each trial, a fixation cross was presented in the centre of the screen (2-5 seconds). Participants were informed that they should make these choices according to their own preferences, as 40 of these 50 trial choices would be used in the scanning session.

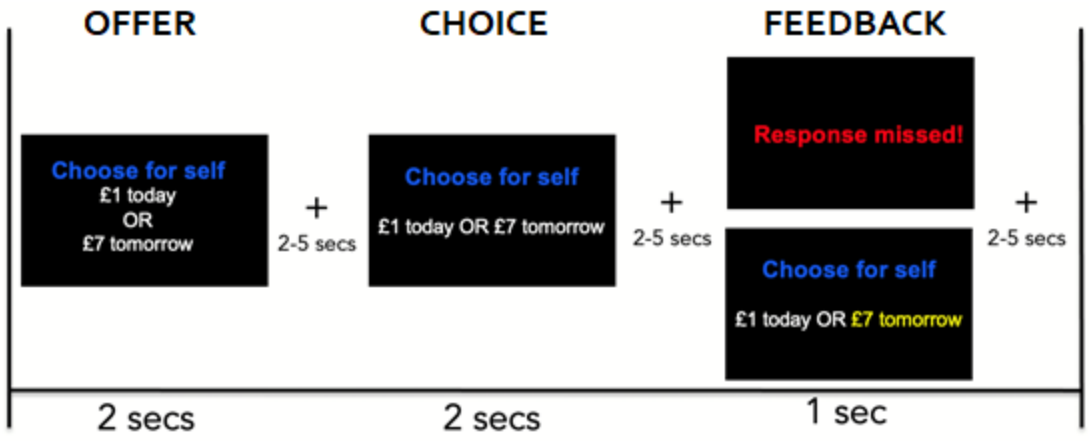


Figure 2. Training task experimental design. This shows the three experimental screens (Offer, Choice and Feedback stages) shown to participants in the training task and the timing of a single trial.

Temporal discounting contagion fMRI task

The TD contagion task presented to participants in the scanner was divided into five blocks of 40 trials (Self1, Other1, Self2, Other2, and Self3). See **Figure 3** for the order of these blocks, and example trials for Self and Other blocks. A ten second break

was included at the start of the task to allow for scanner calibration. At the start of each block, an instruction screen was displayed for four seconds, to inform the participant who they would be making choices for in the following block (e.g., *"Make choices for yourself."*), and participants were given a five second break in the middle of each block (after 20 trials), and an eight second break between blocks. Across all five blocks, the same basic trial structure was implemented following the same format as the training task. See **Supplemental Information** for details of how the choice pairs for each trial were generated.

Figure 3. Task order and experimental screens. This shows the order and type of the five blocks for each task. The Self, Other1 and Other2 text colours were randomised across participants (RGB colours). In all Self-trials, participants saw the Offer screen (options presented on top of each other), and the Choice screen (options side by side). In all Other-trials, participants saw the Offer and Choice screens, as well as an additional feedback screen.

The location of the immediate and delayed options on the screen at the Offer Stage was unrelated to the location of these choices on the screen at the Choice Stage, and as such, participants were unable to use information that appeared on the screen at the Offer Stage to prepare their motor response in the Choice Stage. This allowed us to separate activity related to choice evaluation at the Offer stage from motor activity in the

Choice stage. In order to be able to time-lock blood oxygen level-dependent (BOLD) activity in our task to specific trial events throughout the entire experiment (without contaminating effects of other trial events), and to temporally separate each trial from the previous and following trials, we introduced a variable delay between trials, and between trial screens (2-5 seconds), as in previous studies (e.g., Balsters & Ramnani 2008, 2011). A 2-5 second variable delay was used because this produced fewer correlations between trial events than a 2-4 second delay (i.e., across 2 TRs). With this variable delay, the experiment took between 35 and 55 minutes for participants to complete.

Block 1 (Self1)

During this block, participants were instructed to choose for themselves according to their own preferences. In order to ensure that participants chose according to their own preferences on all Self-trials, participants were informed that one of the choices they made for themselves (across all three Self-blocks) would be randomly selected and used as their bonus payment. In fact, all participants were paid a randomly selected amount between £10 and £15. Participants viewed two experimental screens (i.e., two experimental events) in this block (Offer Stage and Choice Stage), presented in the same format as in the training task.

The Offer was displayed to participants for 2 seconds, followed by a fixation cross (2-5 seconds variable delay). The Choice screen was then displayed, and participants had 2 seconds to execute their choice by pressing the relevant button on an MR-safe button box. The Choice screen remained visible for the full two seconds, whether or not participants executed a response during this time. No feedback was shown to participants in this block.

The timings of the Offer and Choice stages were entered as conditions into the design matrix. The trial-by-trial estimation of participants discounting parameters (log *k* values) were entered as a parametric modulator for the Offer condition, and RTs were entered as a parametric modulator for the Choice condition.

Block 2 (Other1)

In this block, participants were instructed to choose for either the Close Other, or the Distant Other/stranger, with the order counterbalanced across participants. The direction of the Other (i.e., ± 1 of the participant's own discounting parameter, calculated at the end of the Self1-block) was also counterbalanced across participants. A more positive agent is more impulsive than the participant, and a more negative agent is more patient.

Participants viewed three experimental screens in this block (Offer Stage, Choice Stage and Feedback Stage). The Offer and Choice screens were displayed in the same format (with the same timing) as the training task, and Block 1. Following the Choice screen, the fixation cross was shown for between 2 and 5 seconds. The participant then viewed the Feedback screen for 1 second, which informed them whether they had made the correct choice for the other participant (e.g., *"Correct! Your decision for Imogen was correct."* displayed in green font, or *"Wrong. Your decision for Imogen was incorrect."* displayed in red font). Participants believed that the correct choice was the option that was selected by the other participant in the training phase. In fact, feedback was displayed as correct if participants chose the option that was selected by the model.

The event timings of the Offer, Choice and Feedback stages were entered as conditions into the design matrix. The trial-by-trial estimation of participants discounting parameters (log *k* values) when making choices on behalf of the Other were entered as a

parametric modulator for the Offer condition, and RTs were entered as a parametric modulator for the Choice condition. For the Feedback condition, an estimation of the surprise that participants experienced when comparing the choice that they predicted their partners would make, versus the choice their partner actually made was entered as a parametric modulator. We also entered a binary measure of whether or not participants selected the option with the highest subjective value, according to the model used to produce the choices of the other (see **Supplemental Information** for calculation of these two measures).

Blocks 3, 4, and 5

Blocks 3 and 5 (Self2 and Self3) followed the same format as Block 1, with the same variables entered into the design matrix for each of the conditions (e.g., Self2 Offer, with Self2 log *k* as a parametric modulator, Self2 Choice, with RT as a parametric modulator). Block 4 (Other2) followed the same format as block 2, with the same variables entered into the design matrix (i.e., Other2 Offer, with log *k* as a parametric modulator, Other2 Choice, with RT as a parametric modulator, and Other2 Feedback, with the surprise measure, and the binary subjective value measure as parametric modulators). Participants made choices for the remaining other (e.g., if Other1 was a more negative Distant Other, Other2 was a more positive Close Other).

Estimation of discount rates and simulation of the Other's choices

A log hyperbolic model was fitted to participants' choices in the scanning task. Training data was used only for training and familiarising with the task, and thus this data was not analysed. The choices of the two modelled agents were also produced using

this log hyperbolic model. The subjective value of each choice was calculated on each trial according to the following equation:

(1) $$V = \frac{M}{1+kD}$$

Subjective value, and the reward magnitude are represented by V, and M respectively, *k* represents the agent's log *k* value, and *D* indicates the delay period in days. Log *k* parameter values were set between -4 and 0. A log *k* value of -4 indicates that the participant bases their decision on reward value and is not sensitive to delay. Participants with log *k* values closer to 0 discount rewards more steeply as a function of time and are more sensitive to delay (Carlisi et al., 2017; Garvert et al., 2015). The subjective value of the immediate option will always correspond to the magnitude of the reward (*M*), because the delay period in days (*D*) is 0.

In the following softmax function, the subjective value of the immediate option is referred to as V_{SS} and the value of the delayed option is referred to as V_{LL}. This function was used to transform the subjective values of each option into choice probabilities (i.e., P_{SS} and P_{LL}):

(2) $$P_{LL} = \frac{1}{1+e^{-\beta(V_{LL}-V_{SS})}}$$

In this equation, β is a free parameter, which characterises noise in an individual's choices. Log β parameters were set between -1 and 1. Values closer to -1 indicate larger non-systematic deviations around the indifference point (i.e., the point at which both choice options are equally preferred).

The choices of the Close and Distant others were modelled within the task (based on the participant's own choices in the Self1-block) to reflect the preferences of an agent with a log *k* value plus or minus one from the participant's own log *k*. Differences in subjective value of the two options were translated into choice probabilities using

equation (2) with a fixed β of 1. An agent with a larger log *k* than the participant discounts rewards more steeply and tends to prefer immediate options (i.e., more impulsive). An agent with a smaller log *k* than the participant discounts rewards less steeply, and tends to wait longer for rewards (i.e., is more patient).

Participants' own log *k*, and log β values were also derived from equations (1), and (2), and were updated on a trial-by-trial basis using Bayesian statistics, with the log β parameter set to start at 0.3 for Self-blocks. A uniform prior was updated on each trial, and the posterior was calculated as the likelihood of an individual's choice given the parameters log *k*, and log β weighted by this prior. The posterior then became the prior for the next trial. To avoid the potential influence of strong priors for the Close and Distant agents on the estimation of priors for Self, and because previous research has indicated that we use our own priors as a starting point for learning about others (Lockwood et al., 2018; Suzuki et al., 2016; Tarantola et al., 2017), priors were reset at the beginning of each Self-block. The posterior from the end of the Self1-block then became the prior for the start of the Other1-block, and this reset at the start of the Self2-block. The posterior from the end of the Self2-block then became the prior for the Other2-block, and this reset at the start of the Self3-block.

Behavioural contagion when accounting for uncertainty

A Bayesian belief update measure (Kullback-Leibler divergence (D_{KL}); Kullback & Leibler, 1951) was used to calculate the behavioural change in participants' prior beliefs about log *k* values between blocks (i.e., contagion). This measure quantifies the divergence in the distribution of two data sets (i.e., the posterior at the end of each Self-block, after making choices on behalf of the Other agent), and accounts for both the

change in the overall peak of this distribution, and the spread or precision of the distribution, which reflects the participant's certainty in their belief. We normalised D_{KL} for our behavioural analyses of contagion, such that D_{KL} was positive when participant's log *k* changed in the same direction as the Other (i.e., the participant's own log *k* became more positive after making choices on behalf of an agent whose own log *k* was also more positive than the participant's initial log *k*), and D_{KL} is negative when participant's own log *k* changes in the opposite direction to the Other.

Data acquisition and pre-processing

MRI data was collected on a 3 Tesla Siemens TIM Trio MRI scanner with a 32-channel head coil, at Royal Holloway, University of London. Before the main experimental task began in the scanner, we acquired a high-resolution T1-weighted anatomical MPRAGE image (voxel size = 1 mm × 1 mm × 1 mm). Each participant then completed the main experimental task, which was a single EPI session, lasting between 35 and 55 minutes (depending on length of jitter), containing 46 slices (multi-slice mode interleaved, collected in descending order). Scans were acquired with a voxel size of 2.5 mm x 2.5 mm x 2.5 mm, a 2 second TR, 76-degree flip angle, a 34 ms echo time and a field of view covering 100% of the brain.

Scans were pre-processed using SPM12. The anatomical scans were first co-registered to the SPM canonical T1 weighted MPRAGE 1 x 1 x 1 mm voxel image, and then all functional images for each participant were co-registered to this anatomical image. Functional images were then realigned (estimate and reslice) registered to the mean, using SPM defaults (with quality set to 1). These realigned images were then co-

registered with the anatomical image. Images were then spatially normalised (estimate and write) to a standard EPI template using SPM default settings.

fMRI analysis

The first-level model had 12 regressors, and each regressor had one or two parametric modulators (outlined in the **Temporal discounting contagion fMRI task** section above) and was specified and estimated using the Robust Weighted Least Squares (RWLS) toolbox (Diedrichsen & Shadmehr, 2005) in SPM 12. For each participant, we included the six motion regressors obtained during realignment as covariates.

We specified multiple paired contrasts to determine whether we would find an effect of contagion (for Close and Distant Others separately), whether there was a difference in contagion for Close versus Distant Others, and whether there were differences in activation when participants were making choices on behalf of Close versus Distant Others at the Offer, Choice, and Feedback stages. Spatial smoothing was applied to these contrast images using a 5mm full width by half-maximum Gaussian kernel.

For the whole brain multivariate analysis, we created 4D images of all participant contrasts, and conducted 10,000 random permutations of these 4D images using FSL. For all analyses, TFCE-corrected statistics are reported ($p < .05$).

Regions of interest

Seven $10mm^2$ diameter ROIs were selected (see **Figure 4**), with the peak coordinates of each of these identified by Garvert et al. (2015; see **Table 1** and **Table 2**). ROIs were located in the medial prefrontal cortex (mPFC; two ROIs in this area),

posterior medial frontal cortex (pMFC), right insula, left insula, left striatum, and left superior temporal sulcus (STS).

Statistical analyses – behavioural data

Analyses for all three studies are run in log space. All analyses were run using Jamovi (version 1.1.9), JASP (version 0.13.0.0), and R (version 3.5.2) in RStudio (version 1.1.463). G*Power (version 3.1) was used for power calculations.

Results

Belief in manipulation, and social closeness

All participants in the final sample completed the awareness of manipulation (i.e., *"Did you believe that the choices of the other two players were real?"*) and change awareness (i.e., *"Did you notice yourself changing your choices throughout the experiment?"*). A total of 27 participants (90%) believed that they were viewing the choices of their testing partner and the stranger (i.e., confederate). In a previous study (Thomas et al., 2021) only 40.23% of participants believed the manipulation, indicating that using real-world pairs of close others successfully increased belief in the manipulation in comparison to previous research. A total of 23 participants noticed their behaviour changing (76.67%).

To determine whether belief in manipulation or awareness of change affected contagion for Close and Distant Others, independent-samples t-tests were run. Belief in manipulation did not have a significant effect on contagion for Close (t (28) = 1.34, p = .192, d = .81, BF_{01} = 1.22) or Distant (t (28) = .02, p = .985, d = .01, BF_{01} = 2.08) Others, although Bayesian support for the null hypothesis was anecdotal. Awareness of

behavioural change also did not have a significant effect on contagion for Close (t (28) = -.78, p = .442, d = -.34, BF_{01} = 2.08) or Distant (t (28) = .34, p = .738, d = .05, BF_{01} = 2.48) Others.

A paired-sample t-test showed that the difference in participant ratings for Close versus Distant Others was significant (t (29) = 13.34, p < .001, d = 2.44, BF_{10} = 1.058e+11) with higher ratings given for the feeling of social closeness with the Close Other (mean = 4.80, SD $\pm$ 1.56) versus the Distant Other (mean = 1.03, SD $\pm$ 0.18). This indicates that participants did feel closer to their testing partners versus the strangers, providing a good basis for exploring the existence of a social distance effect on contagion and learning.

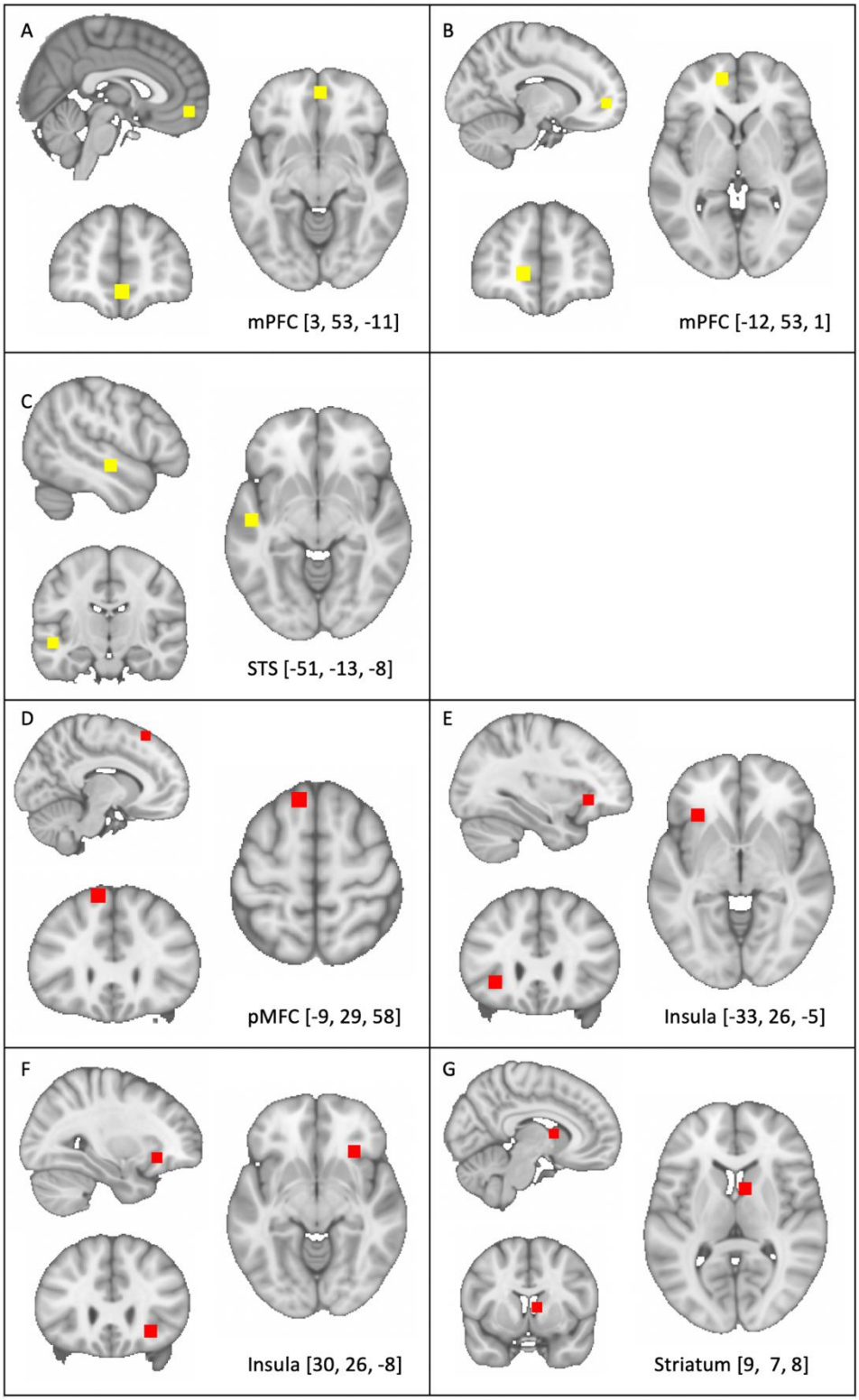


Figure 4. Regions of interest. This shows the seven ROIs used in this paper, identified by Garvert et al (2015). A), B), and C) are presented in yellow, and show the three ROIs associated with contagion in the Garvert paper, and D), E), F), and G) are presented in red, and show the four ROIs associated with learning in the Garvert paper.

Contagion of temporal discounting value preferences

One-sample t-tests revealed that contagion (as measured by D_{KL}) was significantly greater than zero for both Close (t (29) = 5.77, p < .001, d = 1.05, BF_{10} = 6327.57) and Distant (t (29) = 6.47, p < .001, d = 1.18, BF_{10} = 37857.09) Others, indicating that participants were influenced by the choices of both agents. Whilst contagion was slightly higher for Close Others, a paired-sample t-test showed this difference was not significant (t (29) = 1.32, p = .196, d = .24, BF_{01} = 2.34), and a TOST paired-samples t-test (with equivalence bounds set between -.7 and .7) showed significant support for equivalent contagion for Close and Distant Others at the lower (t (29) = -2.51, p = .009) and upper (t (29) = 5.16, p < .001) bounds. See **Figure 5** for a comparison of D_{KL} for Close and Distant Others. These findings show that the degree of behavioural contagion was similar for both Close and Distant Others in our sample.

We extracted parameter estimates (i.e., mean beta values) from three ROIs (i.e., two areas in the mPFC, and one in the STS) identified as being related to contagion of value preferences (Garvert et al., 2015; See **Table 1**). One-sample t-tests run on these beta values revealed no significant neural basis of contagion for Close or Distant Others. Our whole brain analysis also revealed no significant effect of contagion (i.e., Self-Offer before versus Self-Offer after) for Close or Distant Others, and there were also no brain areas that were significantly more activated for the contagion contrast for Close Others versus Distant Others.

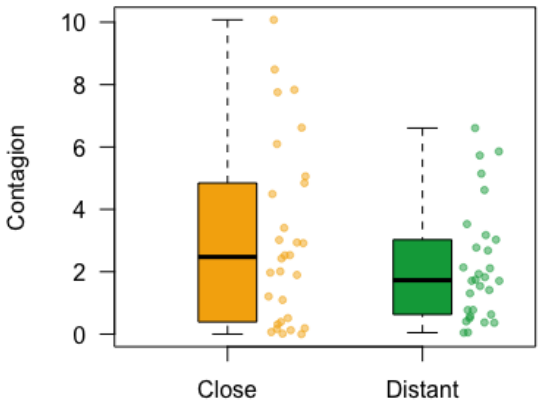


Figure 5. Comparison of D_{KL} for Close and Distant Others.
Descriptive statistics for divergence score (D_{KL}) variables, split by identity of the Other (i.e., Close/Distant).

Table 1. ROIs and statistics for Close and Distant contagion contrasts.

Brain region	L/R	Peak voxel			Close contagion			Distant contagion		
		x	y	Z	*p*	*d*	BF_{01}	*p*	*d*	BF_{01}
mPFC	L	-12	53	1	.515	.12	4.21	.537	-.11	4.30
mPFC	R	3	53	-11	.579	.10	4.45	.179	-.25	2.19
STS	L	-51	-13	-8	.469	-.13	4.02	.182	-.25	2.22

Note: BF_{10} is used for significant analyses, and BF_{01} for non-significant.

Learning the value preferences of other agents

A paired-samples t-test revealed an effect of social distance on accuracy. Participants chose the option preferred by the other agent more frequently for the Close Other than the Distant Other (t (29) = 4.49, $p < .001$, $d = .82$, $BF_{10} = 252.06$, **Figure 6**).

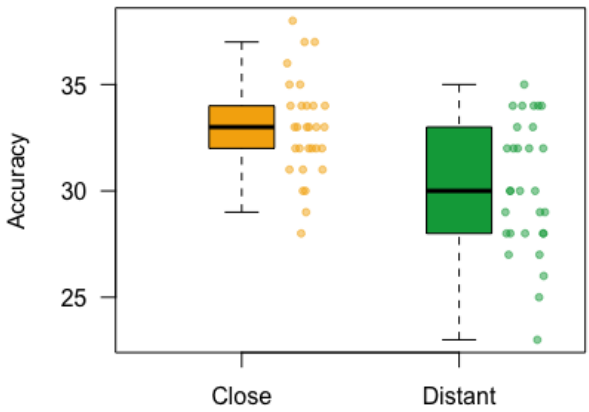


Figure 6. Comparison of accuracy for Close and Distant Others. Descriptive statistics for participant accuracy (out of 40) at making choices on behalf of the other agent, split by identity of the Other (i.e., Close/Distant).

To determine whether there were differences in reaction times (RTs) depending on which agent the participant was making choices on behalf of (i.e., Self, Close, Distant, see **Figure 7** for a plot of RTs with Self averaged, and **Supplemental Figure 1** for a plot of RTs separated for each Self-block), we calculated average RT for Self across the three Self-blocks and compared this with RTs for Close and Distant Others using a repeated-measures ANOVA. The analysis was significant (F (2, 58) = 6.26, p = .003, η^2 = .03, BF_{10} = 10.39), and Bonferroni post-hoc tests showed that the difference in RT was significant for Self versus Close only (p = .002), with longer RTs for Self than for Close. There was no significant difference in RT for Self versus Distant (p = .384), nor Close versus Distant (p = .156).

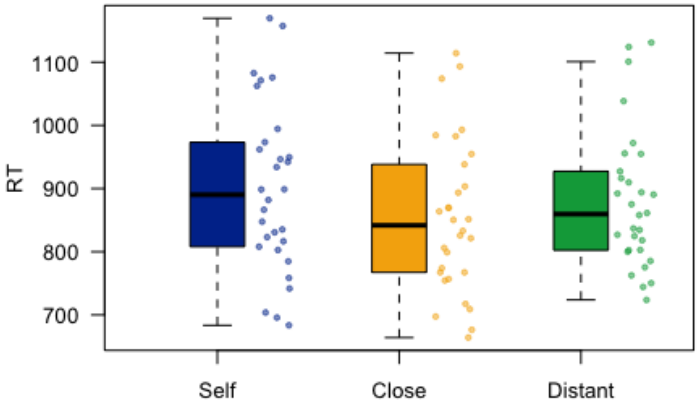


Figure 7. Comparison of RTs for each agent. Descriptive statistics for RT variables, for Self (averaged across blocks), Close other, and Distant Other.

We defined four ROIs in the pMFC, insula, and striatum based on the findings of previous research (Garvert et al., 2015; See **Table 2** and **Figure 4**), and extracted the mean beta values from these areas for contrasts at the Feedback stage (i.e., Feedback, Surprise, and Binary SV). In the left pMFC, activation related to Surprise at the choice of the Other was greater for Close versus Distant Others (with anecdotal Bayesian support). None of the paired contrasts were significant in any of the remaining three

ROIs, although we found significant activation in some of our one-sample contrasts, and these are presented in the **Supplemental Information.**

Our whole brain analysis revealed one cluster of 370 voxels with significantly greater activation for Close versus Distant Others at the Feedback stage in the left postcentral gyrus (maximum peak, t = 2.75, [-46, -24, 61], see **Figure 8** for this cluster, and **Table 3** for all Cluster statistics). As this cluster covers motor areas, we wanted to determine whether this effect could be attributed to residual BOLD signal from motor responses made just before at the choice stage. This activation was not related to differences in RTs between Close and Distant Others (r = .33, p = .073, BF_{01} = .95).

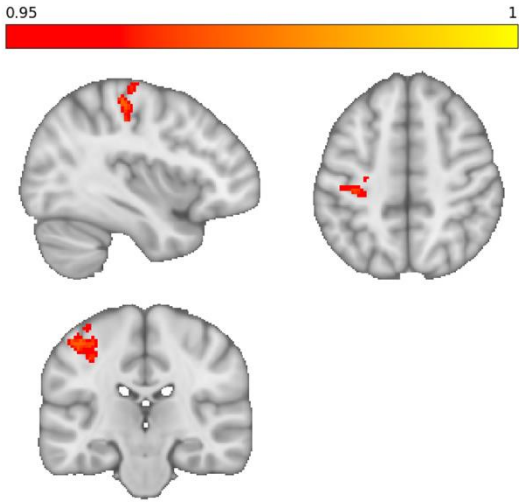


Figure 8. Left postcentral gyrus. This shows the significant Cluster of 370 voxels with greater activation for Close versus Distant Others at the feedback stage, identified in the whole brain analysis (peak [-46, -24, 61], t-stat).

The relationship between learning and contagion

To determine whether behavioural measures of learning and contagion (i.e., D_{KL}) were related, correlational analyses were run between accuracy and contagion for Close and Distant Others. This correlation was not significant for Close (r = -.13, p = .495, BF_{01} = 3.53) or Distant (r = .31, p = .101, BF_{01} = 1.22) Others, indicating no relationship between learning and contagion in this sample.

Table 2. ROIs and statistics for Close and Distant Feedback contrasts (i.e., Feedback, Surprise, and Binary Subjective Value (SV)).

Brain region	**L/R**	**Peak voxel**			**Contrast**	***p*-value**	***d***	***BF***
		x	**y**	**z**				
pMFC	L	-9	29	58	*Other Close > Other Distant Feedback*	.183	.25	2.23
					Other Close > Other Distant Surprise	**.033***	**.41**	**1.67**
					Other Close > Other Distant Binary SV	.055	-.37	.90
Insula	L	-33	26	-5	*Other Close > Other Distant Feedback*	.568	.11	4.41
					Other Close > Other Distant Surprise	.053	.37	.89
					Other Close > Other Distant Binary SV	.835	.04	5.04
Insula	R	30	26	-8	*Other Close > Other Distant Feedback*	.147	.27	1.91
					Other Close > Other Distant Surprise	.074	.34	1.14
					Other Close > Other Distant Binary SV	.490	.13	4.11
Striatum	L	9	7	8	*Other Close > Other Distant Feedback*	.064	.35	1.02
					Other Close > Other Distant Surprise	.335	.18	3.32
					Other Close > Other Distant Binary SV	.830	-.04	5.03

Note: Significant results ($p < .05$) are indicated with an asterisk. BF_{10} is reported for significant analyses, and BF_{01} is reported for non-significant analyses.

Table 3. Clusters, maximum cluster voxels and statistics for the Close - Distant Other Feedback contrast ($p < .05$).

Cluster	k	Max. (SPM area)	L/R	x	y	z	T	Atlas		
								Harvard-Oxford Cortical	Harvard-Oxford Subcortical	Juelich Histological
1	370	Postcentral gyrus	L	-46	-24	61	2.75	61% postcentral gyrus	63% cerebral cortex 23% cerebral white matter	90% GM primary somatosensory cortex BA1 20% GM primary somatosensory cortex BA3b 12% GM primary somatosensory cortex BA2 10% GM primary motor cortex BA4a 6% GM premotor cortex BA6
		Postcentral gyrus	L	-36	-32	59	3.49	55% postcentral gyrus 6% superior parietal lobule 6% precentral gyrus	69% cerebral cortex 30% cerebral white matter	58% GM primary somatosensory cortex BA3b 36% GM primary somatosensory cortex BA2 34% GM primary motor cortex BA4a 29% GM primary motor cortex BA4p 16% GM primary somatosensory cortex BA1 10% WM corticospinal tract 9% GM superior parietal lobule 5L
		Postcentral gyrus	L	-34	-30	51	3.40	30% postcentral gyrus 8% precentral gyrus	57% cerebral white matter 43% cerebral cortex	51% GM primary motor cortex BA4p

								35% GM primary somatosensory cortex BA3a 31% GM primary somatosensory cortex BA3b 30% WM corticospinal tract 19% GM primary somatosensory cortex BA2 10% GM superior parietal lobule 5L 5% GM primary motor cortex BA4a
	Precentral gyrus	L	-38	-26	71	1.85	20% postcentral gyrus 7% precentral gyrus	32% cerebral cortex	-
	-	L	-26	-24	57	1.92	37% precentral gyrus 12% postcentral gyrus	52% cerebral cortex 46% cerebral white matter	77% WM corticospinal tract 22% GM primary motor cortex BA4p 15% GM premotor cortex BA6 14% GM primary motor cortex BA4a 8% GM primary somatosensory cortex BA3b
	Precentral gyrus	L	-30	-24	69	1.40	34% precentral gyrus 27% postcentral gyrus	68% cerebral cortex 24% cerebral white matter	71% GM Premotor cortex BA6 39% WM Corticospinal tract 24% GM Primary motor cortex BA4a

Discussion

Here, we used fMRI and Bayesian modelling to investigate the neural basis of contagion, and learning of the value preferences of others using real-world close and distant other agents (i.e., close friends or partners, and strangers). We aimed to determine whether social distance affected behavioural and neurobiological measures of contagion of TD preferences and learning about the TD preferences of others. It was predicted that contagion would be found for both close and distant others, with stronger contagion effects, and more successful learning for close versus distant others. We also predicted that brain activation related to contagion, and feedback responses would be greater for close versus distant others.

Using real-world pairs of close friends and actual strangers successfully increased belief in the manipulation in comparison to previous research, and participants' rating of closeness were higher for close versus distant others, indicating a good basis for measuring an effect of social distance on contagion. Significant behavioural contagion effects were found for both close and distant others, although there was no significant social distance effect on contagion. We also found no significant BOLD activation related to contagion in the whole brain analysis, nor in any of the regions of interest identified in previous research (Garvert et al., 2015).

However, we did find a social distance effect on feedback processing and learning. Behaviourally, participants were significantly more accurate at choosing on behalf of the close other versus the distant other. The ROI and whole brain analyses also revealed significant BOLD activation related to learning and feedback processing in multiple areas. Together, these findings suggest that, despite a lack of an effect of social distance on contagion, as well as a lack of a neural basis for contagion, participants were more successful at learning about the TD preferences of individuals that they were close to, and several areas were activated in response to feedback for close others.

Contagion of temporal discounting preferences, and social distance

In line with previous research, we found that learning the TD preferences of another agent induced significant changes in the participant's own discount rate (Apps & Ramnani, 2017; Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Thomas et al., 2021). However, our analysis found no significant differences dependent on social distance, and the strength of contagion was statistically similar for both close and distant others. We also did not find any brain areas that varied significantly with contagion, or that showed more activation for contagion for close versus distant others.

These results support previous research that found no behavioural difference in contagion dependent on whether the participant believed they were making choices on behalf of a real participant or not (Garvert et al., 2015; Suzuki et al., 2016; Thomas et al., 2021), and also the finding that mPFC activity did not vary for familiar and novel others (Garvert et al., 2015). Whilst we were successful in increasing participant's belief in the manipulation (i.e., participants believed that they were viewing the choices of the other participants, and not modelled choices) in comparison to previous research (Thomas et al., 2021), it is possible that an effect of social distance on contagion may be specific to risk-seeking in adolescent samples, as observed by Reiter and colleagues (2019).

Despite previous research finding an effect of contagion in areas including the mPFC (Garvert et al., 2015), our analysis did not reveal an effect of contagion in any brain area. Here, we measured contagion as a difference in evaluation of choices before and after making choices on behalf of another individual. The absence of a significant effect can be difficult to interpret, although it is possible that our method of measuring contagion could be responsible for differences in findings between studies. Here, contagion was examined on a trial-by-trial basis, and previous research indicates that trial-by-trial within-subject estimations of subjective value

in a TD task may not be reliable (Fröhner et al., 2019), thus, our form of analysis may have meant an effect of contagion was not observed.

Here, we also used a block design and compared changes in choice evaluation across blocks as our measure of contagion, whereas previous research has used repetition suppression to measure TD contagion (Garvert et al., 2015). Garvert and colleagues (2015) found that representations for all other agents became more similar towards the end of the experiment. If the same response were observed here (i.e., the representations for all agents became more similar towards the end of the experiment), differences in response to choice evaluation during Self-blocks may have reduced throughout the task. Indeed, if this difference were smaller towards the end of the experiment, counterbalancing the order of close and distant others in our task, and examining activation for each of these individually, may have cancelled out variation between early and late blocks. Indeed, previous research indicates that responses to stimuli are reduced when the stimuli are repeated (Baron et al., 2013; Grill-Spector et al., 2006; Henson et al., 2000), and participants in our study repeated the same trial format throughout the task. Comparing groups (with some participants deciding on behalf of close others and some deciding on behalf of distant others) and using the same interleaved trial format as Garvert and colleagues (2015) could further test for a social distance effect on contagion.

However, previous research has found an effect of contagion on PD preferences without using repetition suppression (Suzuki et al., 2016). Suzuki and colleagues (2016) found that contagion of PD preferences was driven via activation in the caudate, whereas updating of the value preferences of the other was reflected in dorsolateral prefrontal cortex. Whilst Suzuki and colleagues (2016) used a PD task, and identified different areas to Garvert and colleagues (2015), it is possible that contagion of TD preferences may be observed in a paradigm that does not rely on repetition suppression. The behavioural effect of contagion observed both here and in our previous study (Thomas et al., 2021) is strong, and has replicated the behavioural

contagion effects found in previous research (e.g., Garvert et al., 2015; Suzuki et al., 2016). It is possible that other methods of analysing or examining the fMRI data would also reveal a neural basis for contagion.

Previous research has examined variations in noise in fMRI data, in the absence of a brain-based effect using traditional general linear models, which rely on differences in mean BOLD signal (Balsters et al., 2013). Balsters and colleagues (2013) found a clear behavioural difference in RTs and accuracy between task conditions in a working memory task, in the absence of a significant BOLD response. However, when examining neural oscillations, and using an ICA approach, the researchers found a strong brain-based effect, with neural oscillations in various functional networks varying in line with working memory task performance. Although the RWLS toolbox was used in the present experiment (which reduces noise in the data), it is possible that the data is noisy, and that variations in signal could also track updates in participants' own subjective values. Future research could explore differences in subjective value update that occur in line with neural oscillations.

Learning the choices of others, and processing feedback

In line with previous research (Behrens et al., 2007; Garvert et al., 2015), we found BOLD responses related to learning, and our findings supported the prediction that we would find a social distance effect on feedback processing. Behaviourally, we found greater accuracy for close versus distant others. This effect also coincided with differences in the BOLD response for close versus distant others. In our ROI analysis, we found significantly greater activation in the pMFC for surprise for the close versus distant other, and a cluster of voxels in the postcentral gyrus that showed greater activation for feedback for close versus distant others. Our behavioural findings also revealed a social distance effect on accuracy, with greater accuracy for close versus distant others.

A significant social distance effect in the somatosensory cortex (i.e., the postcentral gyrus) at the feedback stage may reflect greater activation in response to imagining the close versus distant others receiving feedback, as previous research has implicated this area in predicting the actions of others (Lamm et al., 2007). Research has also shown that empathy can be higher for individuals we are closer to (Bucchioni et al., 2015), more similar to (Majdandžić et al., 2016), or who are part of an in-group (Levy et al., 2016), which may contribute to potential differences in processing feedback for close versus distant others.

The finding of a social distance effect on surprise coding in the pMFC furthers the findings of Garvert and colleagues (2015), who found a PE signal (i.e., the neural representation of the surprise measure) in the pMFC. Here, we show that this PE signal is greater for the choices of those that participants are close to, versus strangers. As PE-type signals in the multiple brain areas are known to drive learning (e.g., Balsters et al., 2017; Behrens et al., 2007; 2008; Garvert et al., 2015; Lee & Harris, 2013; Reynolds & Wickens, 2002), and that these are affected by social distance (Carp et al., 2008; Fareri et al., 2012; Kang et al., 2010; Powers et al., 2016), the larger PEs observed in response to the choices of close versus distant others in this study likely contributed to the increased accuracy (and thus more successful learning) for the TD preferences of close versus distant others.

Consistent with a Bayesian model of social influence, it is possible that differences in activation for close versus distant others here is related to differences in the perceived credibility of social information from close versus distant others. Previous research has indicated that credibility of social information (determined by group size or number of reviews) is reflected in brain areas close to the pMFC (De Martino et al., 2017; Park et al., 2017). Indeed, participants in the present study could have viewed information from close others as being more credible than information from distant others. Further research could vary the credibility of social information in the TD contagion task by incorporating decisions for individuals and

groups (as per Apps & Ramnani, 2017), to determine whether this difference in activation is indeed related to credibility.

Conclusions

Taken together, these findings demonstrate that learning the TD value preferences of both close and distant other agents robustly shifts participants' own value preferences (i.e., contagion). Whilst we did find a behavioural contagion effect for both agents, our results did not support a behavioural or neurobiological effect of social distance on contagion, despite using real-world pairs of close and distant others. An effect of social distance was found on learning and feedback processing, suggesting that participants were more successful at learning about individuals that they were close to. Future research could confirm or refute the absence of a brain-based social-distance contagion effect in TD by testing for an existence of a social distance effect on contagion using an established repetition suppression paradigm (Garvert et al., 2015), or examining whether neural oscillations are related to subjective value updating (Balsters et al., 2013).

References

Apps, M. A. J. & Ramnani, N. (2017). Contributions of the medial prefrontal cortex to social influence in economic decision-making. *Cerebral Cortex, 27*(9), 4635-4648.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self-scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4). 596-612.

Balsters, J. H., Apps, M. A. J., Bolis, D., Lehner, R., Gallagher, L., & Wenderoth, N. (2017). Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain, 140*(1), 235–246.

Balsters, J. H., & Ramnani, N. (2008). Symbolic representations of action in the human cerebellum. *NeuroImage, 43*(2), 388-398.

Balsters, J. H., & Ramnani, N. (2011). Cerebellar plasticity and the automation of first-order rules. *The Journal of neuroscience, 31*(6), 2305–2312.

Balsters, J. H., Robertson, I. H., & Calhoun, V. D. (2013). BOLD frequency power indexes working memory performance. *Frontiers in Human Neuroscience, 7*, 207.

Baron, H. C., Dolan, R. J., & Behrens, T. E. J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience, 16*, 1492-1498.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*, 1214-1221.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*(7219, 245-259.

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron, 62*, 733–743.

Bucchioni, G., Lelard, T., Ahmaidi, S., Godefroy, O., Krystkowiak, P., & Mouras, H. (2015). Do we feel the same empathy for loved and hated peers? *PLoS ONE, 10*(5), 1-11.

Carlisi, C. O., Norman, L., Murphy, C. M. Christakou, A., Chantiluke, K., Giampietro, V., … Rubia, K. (2017). Comparison of neural substrates of temporal discounting between youth with autism spectrum disorder and with obsessive-compulsive disorder. *Psychological Medicine, 47*(14), 2513-2527.

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2008). Perceived similarity and neural mirroring: Evidence from vicarious error processing. *Social Neuroscience, 4*(1), 85-96.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591-621.

Courtney, A. L., & Meyer, M. L. (2020). Self-other representation in the social brain reflects social connection. *The Journal of Neuroscience, 40*(29), 5616-5627.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgements according to its reliability. *The Journal of Neuroscience, 37*(25), 6066-6074.

Diedrichsen, J., & Shadmehr, R. (2005). Detecting and adjusting for artefacts in fMRI time series data. *NeuroImage, 27*(3), 624-634.

Fareri, D. S., Niznikiewicz, M. A., Lee. V. K., & Delgado, M. R. (2012). Social network modulation of reward-related signals. *Journal of Neuroscience, 32*(26), 9045-9052.

Fröhner, J. H., Teckentrup, V., Smolka, M. N., & Kroemer, N. B. (2019). Addressing the reliability fallacy in fMRI: Similar group effects may arise from unreliable individual effects. *NeuroImage, 195*, 174–189.

Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J., & Dolan, R. J. (2015). Learning-induced plasticity in medial prefrontal cortex predicts preference malleability.

Neuron, 85(2), 418–428.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Neuroscience, 10*(1), 14-23.

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise signals in anterior cingulate cortex: Neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *Journal of Neuroscience. 31*, 4178–4187.

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science, 287*(5456), 1269-1272.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value guided choice. *Nature Neuroscience, 15*, 470–476.

Izuma, K. (2013). The neural basis of social influence and attitude change. *Current Opinion in Neurobiology, 23*(3), 456-462.

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *PNAS, 105*, 4507–4512.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10,* 1625–1633.

Kang, S. K., Hirsh, J. B., & Chasteen, A. L. (2010). Your mistakes are mine: Self-other overlap predicts neural response to observed errors. *Journal of Experimental Social Psychology, 46*(1), 229-232.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79-86.

Lamm, C., Fischer, M. H., & Decety, J. (2007). Predicting the actions of others taps into one's own somatosensory representations--a functional MRI study. *Neuropsychologia, 45*(11), 2480–2491.

Levy, K., Goldstein, A., Influs, M., Masalha, S., Zagoory-Sharon, O., & Feldman, R. (2016). Adolescents growing up amidst intractable conflict attenuate brain response to pain of outgroup. *PNAS, 113*(48), 13696-13701.

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications, 9*(4747), 1-11.

Majdandžić, J., Amashaufer, S., Hummer, A., Windischberger, C., & Lamm, C. (2016). The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. *Cognition, 157*, 24-38.

Meyer, M. L., Zhao, Z., & Tamir, D. I. (2019). Simulating other people changes the self. *Journal of Experimental Psychology: General, 148*(11), 1898-1913.

Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of inter-temporal preferences. *PLoS Computational Biology, 12*(7), 1-17.

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior, 64*(3), 263-276.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114-1121.

Powers, K. E., Somerville, L. H., Kelley, W. M., & Heatherton, T. F. (2016). Striatal associative learning signals are tuned to in-groups. *Journal of Cognitive Neuroscience, 28*(9), 1243-1254.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences 13*(10), 420-428.

Reiter, A. M. F., Suzuki, S., O'Doherty, J. P., Li, S-C., & Eppinger, B. (2019). Risk contagion by peers affects learning and decision-making in adolescents. *Journal of Experimental Psychology: General, 148*(9), 1494–1504.

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., & Gvirts, H. Z. (2019). Herding brains: A core neural mechanism for social alignment. *Trends in Cognitive Sciences, 23*(3), 174-186.

Suzuki, S., Jensen, E. L. S., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioural contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *PNAS, 113*(14), 3755-3760.

Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications, 8*(817), 1-14.

Thomas, L., Lockwood, P. L., Garvert, M. M., & Balsters, J. H. (2021). Contagion of temporal discounting value preferences in neurotypical and autistic adults. *Journal of Autism and Developmental Disorders.* Advance online publication.

Tunçgenç, B., El Zein, M., Sulik, J., Newson, M., Zhao, Y., Dezecache, G., & Deroy, O. (2021). Social influence matters: We follow pandemic guidelines most when our close circle does. *British Journal of Psychology.* Advance online publication.

Supplemental Information

Supplemental methods

Generation of choice pairs

In the three Self-blocks, choice pairs for all trials were generated according to two methods (i.e., generative and adaptive), which were alternated across trials (as per Garvert et al., 2015). Two methods were used here in order to ensure accurate estimations of participants' discount rates. The generative method involved generating all possible pairs of amounts and delays, and 20 trials per block were selected that best matched the indifference points of 20 hypothetical participants with log *k* values evenly distributed between -4 and 0 (Garvert et al., 2015). This method was also used to generate all 40 Other-trial choice pairs for the two Other-blocks.

The remaining 20 trials for each Self-block were generated according to an adaptive method, which relies on a Bayesian framework to produce precise estimations of discounting parameters. The participant's initial prior belief about log *k* (at the start of the Self1-block) was set to be normally distributed, with a mean of -2 and a standard deviation of 1. After each decision, Bayes rule was used to form the posterior by updating this prior. Choice pairs were generated which probed participant's own indifference points, in order to refine the prior.

Surprise and binary subjective value measures

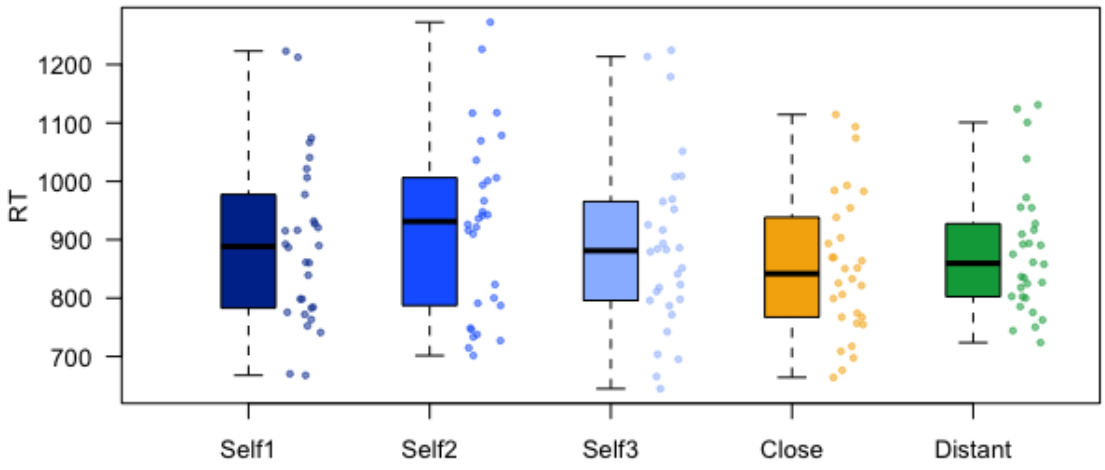
To calculate this measure, we used trial-by-trial estimates of discount rate (when participants were making choices on behalf of the other) and used these choices to compute differences in subjective value between the choices they predicted their partner would make (i.e., the choice that they selected themselves for the other). The difference in subjective value was then transformed into a probability using a softmax function applied to trial-by-trial estimates of the participant's own log β values. This measure gave an estimate of how likely

the participant would have been to have made the same choice for themselves. This measure was then subtracted from 1 to translate to a surprise measure. This was calculated following the same procedure employed by Garvert and colleagues (2015).

A binary subjective value measure (i.e., 1 if the participant choice the option that would have the highest subjective value for the other agent, and 0 if they chose the other option) was also included because we introduced noise into model which produced the choices of the other participants, and the chosen option did not always reflect the option with the highest subjective value according to this model. This measure, therefore, acted as a measure of how well the participant was tracking the preferences of the model, instead of an outright correct or incorrect response, which would show how well the participant was tracking the noise in the model.

Supplemental analyses

Learning the value preferences of other agents



Supplemental Figure 1. Comparison of RTs for each agent. Descriptive statistics for RT variables, for Self1, Self2, Self3, Close other, and Distant Other.

One-sample brain imaging contrast results – Regions of interest

We defined four ROIs in the pMFC, insula, and striatum based on the findings of previous research (Garvert et al., 2015), and extracted the mean beta values from these areas

for one-sample contrasts at the Feedback stage (i.e., Feedback, Surprise, and Binary SV). See **Supplemental Table 1** for full details of these contrasts.

A one-sample contrast revealed significant activation in the pMFC for Close Other Feedback and Close Other Surprise. In the left insula, significant activation was found in the Surprise condition for Close and Distant Others, with extreme Bayesian support for Close and very strong Bayesian support for Distant Others. Right insula activity was significant Close Others in the Feedback and Surprise conditions, with strong and extreme Bayesian support respectively. Activation in the right insula for Distant Others was also significant in the Surprise and Binary SV conditions, although Bayesian support for these findings was anecdotal.

In the left striatum, activation in the Close Other Feedback condition was significant, with very strong Bayesian support. Activation was also significant in the Surprise condition for both Close and Distant Others, with strong Bayesian support for the finding for the close other and anecdotal support for the finding for the Distant Other.

Supplemental table 1. ROIs and statistics for Close and Distant feedback one-sample contrasts.

Brain region	L/R	Peak voxel			Contrast	*p*	*d*	*BF*
		x	y	z				
pMFC	L	-9	29	58	***Close Other Feedback***	**.004***	**.57**	**10.18**
					Distant Other Feedback	.055	.37	.90
					Close Other Surprise	**.005***	**.56**	**8.60**
					Distant Other Surprise	.559	.11	4.38
					Close Other Binary SV	.112	-.30	1.56
					Distant Other Binary SV	.415	.15	3.76
Insula	L	-33	26	-5	*Close Other Feedback*	.059	.36	.95
					Distant Other Feedback	.273	.20	2.92
					Close Other Surprise	**< .001***	**1.00**	**2895.24**
					Distant Other Surprise	**.001***	**.67**	**32.78**
					Close Other Binary SV	.543	-.11	4.32
					Distant Other Binary SV	.277	-.20	2.95
Insula	R	30	26	-8	***Close Other Feedback***	**.001***	**.66**	**27.97**
					Distant Other Feedback	.083	.33	1.24
					Close Other Surprise	**< .001***	**.89**	**691.81**
					Distant Other Surprise	**.043***	**.39**	**1.35**
					Close Other Binary SV	.416	-.15	3.77

					Distant Other Binary SV	**.025***	**-.43**	**2.11**
Striatum	L	9	7	8	***Close Other Feedback***	**<.001***	**.68**	**37.01**
					Distant Other Feedback	.590	.10	4.48
					Close Other Surprise	**.003***	**.59**	**12.17**
					Distant Other Surprise	**.032***	**.41**	**1.70**
					Close Other Binary SV	.298	-.19	3.09
					Distant Other Binary SV	.283	-.20	2.98

Note: Significant results ($p < .05$) are indicated with an asterisk. BF_{10} is reported for significant analyses, and BF_{01} is reported for non-significant analyses

Chapter 6.

Temporal Discounting Contagion and Probabilistic Social Learning: Is Contagion Linked to Advice-Taking?

Louisa Thomas

Joshua H. Balsters

In prep

Chapter starts on **page 56** of the thesis

33 pages in chapter, including references

Abstract

Previous research has found that learning the preferences of other individuals or groups can influence our behaviour, and that this effect may be driven by learning about the other agents. Here, we fit Bayesian computational models to a temporal discounting task, where participants made decisions for themselves before and after learning the alternative value preferences of another individual, and aimed to determine whether contagion on this task was associated with advice-taking and choice behaviour in a probabilistic social reward-learning task. In an online sample (N = 195), we found a strongly significant contagion effect, which supported the findings of previous research. However, contagion in this task was unrelated to both task accuracy, as well as advice-taking and choice behaviour in the probabilistic social reward learning task. Due to the potential impact of the COVID-19 pandemic on levels of anxiety and depression, we also ran exploratory analyses to determine the relationship between contagion, accuracy, and advice taking with measures of anxiety and depression, and found no significant relationships. Together, these results indicate that contagion is a reproducible effect, although the relationship between contagion and behavioural measures of learning is unreliable. Future research should aim to explore associations between task-related activation in brain imaging studies.

Introduction

Multiple previous studies have found that the preferences of other's impacts upon our own preferences in economic tasks, with this change occurring after participants successfully learn about the alternative value preferences of other individuals, or groups (Apps & Ramnani, 2017; Garvert et al., 2015; Nicolle et al., 2012; Reiter et al., 2019; Suzuki et al., 2016; Thomas et al., 2021). Despite the role of learning in social influence effects such as contagion (an implicit social influence effect, whereby learning about other's makes you more like them; Wheeler, 1966), little research has examined whether the ability to track the behaviour of others is related to contagion across different tasks. Here, the relationship between contagion in a temporal discounting (TD) task and advice-taking in a social probabilistic reward learning task is explored.

TD refers to the decrease in subjective valuation of rewards as a function of increasing time. Tasks assessing TD preferences involve a series of inter-temporal choices between smaller immediate rewards, and larger later rewards (Basile & Toplak, 2015). The commonly observed response pattern is that the rate at which rewards are devalued as a function of time slows down as the delay period increases (Ainslie, 1975), and TD data can be well-explained by a hyperbolic model (Garvert et al., 2015; McKerchar et al., 2009). Several behavioural and neuroimaging studies have found contagion of TD preferences (Apps & Ramnani, 2017; Nicolle et al., 2012; Thomas et al., 2019). In these studies, participants were tasked with making TD choices on behalf of themselves, and on behalf of other agents with alternative value preferences. In this study, the choices of the other agents will be modelled to vary according to participant's own preferences, which accounts for variability in participant's own rates of TD (as per Garvert et al., 2015; Thomas et al., 2021).

Despite the role of learning about the behaviour of others in contagion in value preferences (Garvert et al., 2015; Reiter et al., 2019; Suzuki et al., 2016; Thomas et al., 2021), the relationship between contagion in discounting tasks and tracking of the behaviour of others in other decision-making paradigms has yet to be extensively explored. Previous research has used Bayesian belief updating to measure contagion (Garvert et al., 2015; Thomas et al., 2021), as research indicates that Bayesian associative processes underlie learning about both social and non-social information (Behrens et al., 2008; Bennett, 2015; Campbell-Meiklejohn et al., 2010; Perreault et al., 2012; Suzuki et al., 2012). Applied to contagion, Bayesian belief updating accounts for both the overall change in participants' preferences after making choices on behalf of other agents, as well as their degree of certainty in their own beliefs, and the likelihood of social information (Behrens et al., 2008; De Martino et al., 2017; Moutoussis et al., 2016; Park et al., 2017; Vilares & Kording, 2011). Here, Bayesian belief updating is used to measure contagion, and this measure is related to advice-taking and the extent to which participant choices are guided by different types of trial information in a task which is underpinned by Bayesian associative learning processes (Behrens et al., 2008).

The probabilistic reward learning task used here is a two-forced choice card game with associated reward probabilities that has an additional social component (i.e., helpful or misleading social advice indicating which card to choose). This task requires participants to learn from both social and non-social information, and the reliability of the social and reward information (i.e., whether the reward probability and probability of the social information being correct is stable, or volatile) vary independently of each other across the task, allowing for independent learning of both types of information (Behrens et al., 2008; Henco et al., 2020; Sevgi et al., 2020). Autism and psychiatric

disorders have each been associated with differences in Bayesian inference relative to NT participants (e.g., Henco et al., 2020; Karvelis et al., 2018; Lawson et al., 2017), and versions of this task have previously been used to explore tracking of social behaviour in neurotypical (NT) participants (Behrens et al., 2007; 2008; Henco et al., 2020) participants with high levels of autistic traits (Sevgi et al., 2020), and participants with psychiatric disorders (Henco et al., 2020). Differences in task performance have been attributed to differences in the weighting of information, particularly when probabilities associated with reward and social information are volatile (Henco et al., 2020; Sevgi et al., 2020). As accurate updating of beliefs is required for both contagion in TD tasks, and tracking the reliability of social information in the probability task, it is of interest to determine whether advice-taking (which represents following of social information) across stable and volatile phases in the probability task is associated with contagion in the TD task. If participants choices are more closely associated with social information in the task versus non-social information (i.e., reward probabilities), this would suggest a social basis for the contagion effect.

As this study was carried out during the COVID-19 pandemic, we also wanted to account for potential differences in participant self-reported anxiety and depression. Previous disease outbreaks have induced anxiety and fear related behaviours in general populations (Person et al., 2004; Shultz et al., 2016), and research has also shown anxiety and depression in response to the current COVID-19 outbreak (e.g., Rajkumar, 2020; Wang et al., 2020). Whilst little research has explored the impact of self-reported anxiety and depression on social learning and social influence in non-clinical populations, research indicates a difference in sensitivity to social and non-social rewards, as well as a reduced motivation to engage in social interaction in psychiatric disorders including

major depressive disorder (Henco et al., 2020; also see Kupferberg et al., 2016 and Whitton et al., 2015 for reviews). One of the few studies to examine relationships between social learning and trait measures of anxiety in the general population also found impaired learning of the probabilities associated with social punishment and reward under conditions of volatility in those with high social anxiety (Beltzer et al., 2019). Potential differences in the learning of social information dependent on levels of anxiety and depression may therefore impact on contagion and choice behaviour in the card game in this study.

We conducted one online study to examine: (1) whether an effect of contagion of TD preferences can be found in an online sample, and (2) whether contagion of TD value preferences is more closely associated with environmental learning (i.e., learning about the reward probabilities associated with the card) or social learning (i.e., learning about the reward probabilities associated with the social information). We also conducted exploratory analyses to determine whether anxiety and depression impact upon learning and contagion in the TD and choice behaviour in the card game. Participants completed a TD contagion task, a two-forced choice social probabilistic reward learning card game task, and a questionnaire assessing current anxiety and depression. We predicted that we would observe an effect of contagion, in that participants own TD preferences became more like that of the other agent, and that contagion would be related to advice-taking in both stable and volatile conditions.

Methods

Participants

A sample of 195 participants aged between 18 and 30 (80 female, mean age = 22.96, SD ± 3.59) were recruited using Prolific, and were paid £3 for taking part. Seven participants failed attention checks, and were excluded from the analyses (final sample 77 female, mean age = 22.98, SD ± 3.60). Informed consent was obtained, and ethical approval was granted by the Royal Holloway Departmental Ethics Committee.

Procedure

Participants completed two decision-making tasks (one TD contagion task and probabilistic reward learning task presented as a card game), a short attention check, and the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983). Participants first completed one of the decision-making tasks, with the order counterbalanced across participants. This was followed by the HADS, which is a 14-item questionnaire with Likert-scale responses (scored from 0-3), with 6 items used to measure anxiety (HADS-A) and 6 items used to measure depression (HADS-D). Participants are asked to respond to each question by selecting the answer that best matches how they have been feeling in the past week. Scores for each subscale ranging between 0-7 indicate normal, 8-10 borderline abnormal, and 11-21 abnormal (i.e., a case of anxiety or depression). If participants did not respond on a particular question, they were given a score of zero for that question. No participants skipped more than one question on the questionnaire. An attention check was also built into this questionnaire, and participants were instructed to respond, *"Strongly agree"* to this question. Following the questionnaire, another short attention check was presented (with two trials) in which

participants were asked to select a picture of a cat amongst three pictures of dogs. Participants then completed the remaining decision-making task, followed by some final questions.

Temporal discounting contagion task

Participants completed a TD contagion task, which was presented and programmed in Gorilla, and was developed based on the experimental script programmed by Mona Garvert (Garvert et al., 2015). In this task, participants made a series of self-paced hypothetical choices between two monetary values (£1-20) presented simultaneously, choosing between a small amount of money available immediately (e.g. £3 now), or a larger amount available after a specified delay period (i.e., tomorrow, 1 week, 2 weeks, 4 weeks, 6 weeks, 2 months, or 3 months), with the side of the immediate option (i.e., left or right) randomised on a trial-by-trial basis. Participants responded on each trial by clicking on the preferred option.

The task was divided into three blocks of 50 trials (i.e., Self1, Other, Self2,). During Self-trials, participants were instructed to choose for themselves according to their own preferences. In block two participants made choices on behalf of a simulated agent, whose behaviour was modelled within the experimental script to reflect the preferences of an agent whose discount rate parameter (i.e., log k) was +1 or -1 from the participant´s own discount rate. An agent with a discount rate that is +1 that of the participant discounts rewards more steeply (i.e., is more impulsive), and an agent with a discount rate that is -1 that of the participant tends to wait longer for rewards (i.e., is more patient). As significant contagion effects have been found for both more impulsive and more patient agents in TD tasks (Thomas et al., 2021), the order is counterbalanced in

the present study, such that half of the participants made choices on behalf of a more positive agent, and half made choices on behalf of a more negative agent. Whilst the preferences of the other agent were actually modelled in this task, participants were informed that they had been paired with a participant who had previously completed the task, and that they would be making choices on behalf of this participant, with the aim of accurately predicting this participant's preferences over time.

On each trial, differences in value of the immediate and delayed options were translated into choice probabilities (i.e., the probability of selecting each option) using a softmax function with a temperature parameter (log β) fixed at 1. During Other-trials, feedback was displayed on a trial-by-trial basis, to inform participants whether their choice for the other agent was correct. If participants chose the option that would be preferred by the modelled agent, feedback was displayed in the form of a tick indicating that the choice was correct. A cross indicating that the choice was incorrect was shown if the participant chose the other option. An example experimental screen is shown in **Figure 1**.

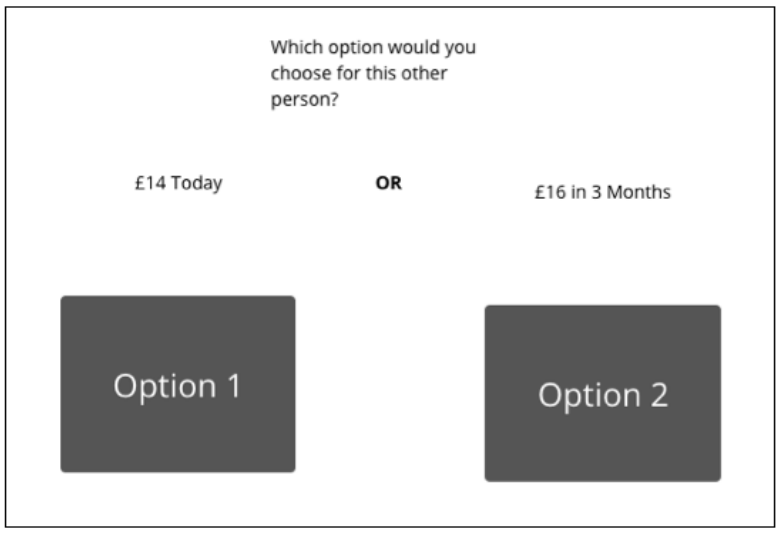


Figure 1. Temporal discounting contagion task experimental screen. This shows an example experimental screen for an Other-trial. Participants saw the two options, along with a question at the top, indicating who they were choosing for (i.e., *"What option would you choose for this other person?"* for Other-trials, or *"Which option would you prefer?"* for self-trials). The items remained on the screen until the participant had selected their response, before a tick or cross (i.e., feedback) appeared in between the two choice options. Following the feedback, the next trial appeared immediately.

Estimation of discount rates and simulation of the Other's choices

In line with previous research (Garvert et al., 2015; Thomas et al., 2021), a log hyperbolic model was fitted to participant's choices, and the choices of the other agent were also produced using this model. Analyses were performed in log space, and log *k* parameter values were set between -4 and 0. The subjective value of each option (i.e., immediate and delayed) was calculated on each trial according to the following equation:

$$(1) \qquad V = \frac{M}{1+kD}$$

V represents the subjective value, *M* represents reward magnitude (i.e., objective value, £1-20), *k* represents the agent's discount rate, and *D* indicates the delay period in days. The subjective value of the immediate reward will always correspond to the reward magnitude (*M*) because the delay period in days (*D*) is 0.

Equation (2) is a softmax function which transformed the subjective values of each option into choice probabilities. The value of the immediate option is referred to as V_{SS}, and the value of the delayed option as V_{LL}, hence the probability of choosing the immediate reward is P_{SS} and the probability of choosing the delayed reward is P_{LL}.

$$(2) \qquad P_{LL} = \frac{1}{1+e^{-\beta(V_{LL}-V_{SS})}}$$

β is a free parameter, which characterises noise in an individual's choices. Log β parameters were set between -1 and 1. The log β parameter was set to start at 0.3 for the two Self-blocks and was fixed at 1 for the Other-block.

Participants' own log *k*, and log β values were also derived from equations (1), and (2), and were updated on a trial-by-trial basis using Bayesian statistics. A uniform prior was updated on a trial-by-trial basis, with the posterior calculated as the likelihood of an individual selecting an option, given the parameters log *k*, and log β weighted by the prior. This posterior then became the prior for the next trial. In line with previous

research (Thomas et al., 2021), the posterior from the end of the Self1-block became the prior for the start of the Other-block, and priors were reset again at the start of the Self2-block. This was to avoid the influence of strong priors for the Other agent on estimation of priors for Self, which could bias the estimate of contagion.

Generation of choice pairs

Choice pairs for all Self-trials (i.e., blocks 1 and 3) were generated according to two methods: generative and adaptive, alternated across trials (as per Garvert et al., 2015 and Thomas et al., 2021). For the generative method, all possible pairs of amounts and delays were generated, and 25 trials per block were selected that best matched the indifference points of 25 hypothetical participants with log *k* values evenly distributed between -4 and 0 (Garvert et al., 2015; Nicolle et al., 2012; Thomas et al., 2021). All 50 Other-trial choice pairs were also generated using this method.

The remaining 25 trials for the two Self-blocks were generated using a Bayesian adaptive staircasing method. The participant's prior belief about log *k* (at the start of the first block) was set to be normally distributed, with a mean of -2 and a standard deviation of 1. After each trial, Bayes rule was used to form the posterior by updating this prior. Choice pairs were then generated in order to probe the participant's indifference point and refine this prior.

Contagion when accounting for uncertainty

Kullback-Leibler divergence (D_{KL}), a Bayesian belief update measure was used to measure contagion (i.e., the change in participants' log *k* values after learning about the preferences of the other agent; Kullback & Leibler, 1951). Contagion can be

interpreted within a Bayesian framework as the integration of an individual's prior beliefs (e.g., about their own discount rate) with their beliefs about newly presented social information (e.g., the discount rate of another individual), in order to form a new posterior belief. The D_{KL} measure calculates the change from prior to posterior belief (i.e., log *k* value before making choices on behalf of the other, versus after), and accounts for both the difference in the position of the prior versus posterior distributions, and the spread (or precision) of each distribution. The change in position corresponds to a straightforward measure of change in log *k* value, with one value subtracted from another (i.e., Self2 log *k* – Self1 log *k*), and the precision reflects an individual's confidence (or certainty) in their belief, as well as the likelihood of the social information (Moutoussis et al., 2016). Here, the end trials of each block are used for calculation, and D_{KL} is normalised, such that changes in log *k* in the same direction as the other agent (e.g., the participant's log *k* became more positive after making choices on behalf of a more positive agent) result in positive D_{KL} values, and changes in log *k* in the opposite direction result in negative D_{KL} values.

Social probabilistic reward learning task format

The task used here is presented as a two forced-choice card game with associated reward probabilities, combined with a social cue (Behrens et al., 2008; Henco et al., 2020; Sevgi et al., 2020). In this task, participants are able to learn from both social and non-social information, and the probability of both the social information being correct on each given trial, and the probability of the blue card (instead of the red card) being the correct card on each trial, vary independently of each other.

Participants completed 120 trials, and on each trial two cards (one red and one blue) with associated reward probabilities were displayed on the screen, with the position of each card (i.e., left or right) determined randomly. Each card had a number on which indicated the number of points associated with the card. These numbers were chosen at random, and totalled to 100 across the two cards. Importantly, the reward magnitudes were independent of the probabilities associated with each card, and thus were not informative of which card would be rewarded on each trial.

On each trial, either the blue or red card was the winning card. If the participants selected the winning card, a star animation appeared out of this card, and the number of points on the card was transferred to their points balance in the top right-hand corner. If participants did not select the winning card, an explosion animation and a red cross appeared over the chosen card. On each trial, one of the cards was also presented in a speech bubble in the bottom right hand corner (i.e., social cue). Participants were informed that this card had been recommended by a previous participant who had completed the game, and knew what the winning card was on each trial. Participants were also informed that the other agent was now playing a new game, where it was sometimes in their best interest to be collaborative (i.e., to provide helpful advice, and tell the participant the correct answer), and sometimes competitive (i.e., to provide misleading advice, and tell the participant the wrong answer for their own benefit in the new game). In fact, the options selected by this other participant were pre-determined. Example experimental screens are shown in **Figure 2**. Whilst some previous research has not provided information about the purpose of the social information (e.g., Henco et al., 2020; Sevgi et al., 2020), explicit feedback about the purpose of the social information was presented here (as per the original study conducted by Behrens et al.,

2008), so that learning occurred in the same manner as in the TD contagion task (i.e., through explicit feedback).

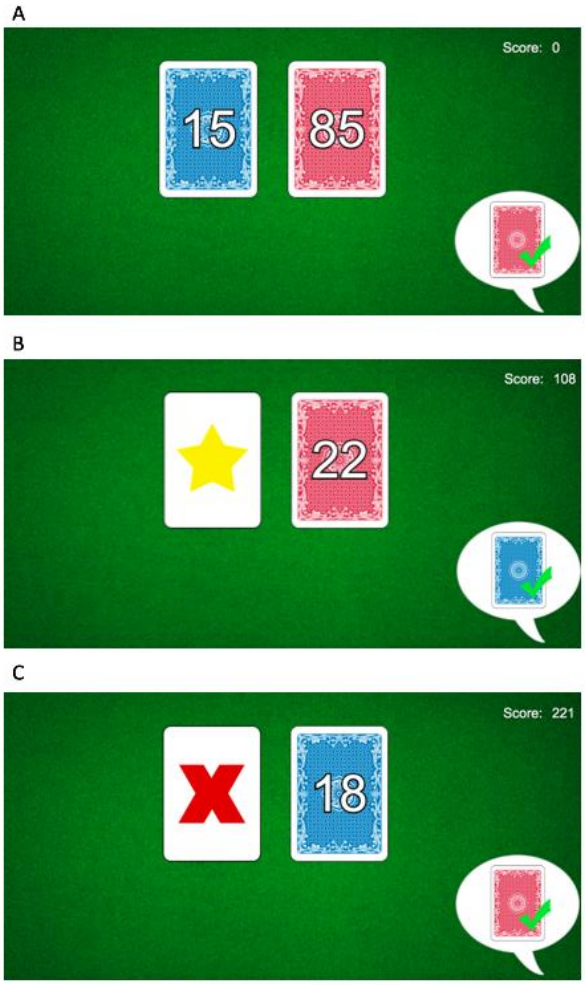


Figure 2. Card game experimental screen. A) Shows an example experimental screen with reward magnitudes displayed on each card, the recommended card displayed in the bottom right-hand corner, and the score (i.e., number of points collected throughout the task) in the top right-hand corner. B) Shows an example trial after the participant made the correct choice. C) Shows an example trial after the participant made the incorrect choice.

Reward and social advice probability schedules

Both the probability of the blue card being the rewarded card, and the probability of the social cue being correct were manipulated throughout the task in line with two independent probability schedules. Manipulating these schedules independently allows participants to learn separately about each source of information (i.e., whether the social information is correct, and which card is most likely to be the rewarded card). There was one reward schedule used for all participants for the probability of the blue card being

the rewarded card (i.e., card probability schedule), and two probability schedules for the probability of the social information being correct (i.e., social information probability schedules). These probability schedules were counterbalanced across participants, such that half of the participants started off seeing helpful social cues, and half of the participants started off as seeing misleading social cues. These conditions are henceforth referred to as the helpful first and misleading first advice conditions. For each of these schedules, there were volatile and stable phases. The probability schedules used here were also used in previous research by Sevgi et al. (2010).

Schedule 1 – Helpful first condition

The probability of the blue card being the winning card was stable and high (i.e., the blue card had a 75% probability of being the rewarded card on each trial) for the first 60 trials of the task. For the second half of the task, the probability of the blue card being the winning card was volatile and changed every 20 trials (i.e., shifted between having a 20% or 80% probability of being the rewarded card).

For half of the participants the advice started out as helpful and was congruent with the rewarded card for the first 30 trials, with the probability of the social cue being correct set at 73% on each of these trials. During the volatile phase (trials 31-70), the probability of the social cue being correct was volatile, and switched between 20% and 80% every 10 trials (starting with 20%). For trials 71-120, the probability of the advice being correct was stable and misleading, with the probability fixed at 16% on each trial.

Schedule 2 – Misleading first condition

The reward probability schedule (i.e., the probability of the blue card being correct) followed the same schedule as in the helpful advice condition, with a stable phase

for the first 60 trials (with a 75% probability of the blue card being the rewarded card), and a volatile phase for the final 60 trials (with the probability changing between 20% and 80% every 20 trials).

For the remaining half of the participants, the probabilities were switched, and the advice started out as misleading and was incongruent with the rewarded for the first 30 trials, with a 27% chance of the social cue being correct on trials 1-30. Between trials 31-70, the social advice was volatile, and switched between having an 80% or 20% chance of being correct every 10 trials (starting with 80%). For the final 50 trials, the probability of the social cue being correct was stable and high, with an 84% chance of being correct on each trial.

Social probabilistic reward task variables

Four advice-taking variables were calculated from the card game data for each participant: (1) the percentage of trials on which the participant followed the advice of the other agent (i.e., chose the card that was displayed in the bottom right hand corner) across all 120 trials, (2) the percentage of trials on which the participant followed the advice in the volatile stage (i.e., trials 31-70 for both conditions), (3) the percentage of advice taken in the stable and helpful stage (i.e., trials 1-30 for the helpful advice condition, and trials 71-120 for the misleading advice condition), and (4) the percentage of advice taken in the stable and misleading stage (i.e., trials 71-120 for the helpful first condition, and 1-30 for the misleading first condition).

For each participant, a regression analysis was also run, and the beta values extracted from this analysis were used as variables in the analyses presented in the results section. For each regression analysis, the participant's choice (i.e., blue =1, or red = 0)

was entered as the outcome variable). Condition 1 was entered as the item with the highest utility (i.e., blue had the highest reward utility = 1, or red = 0). Condition 2 was the card chosen by the social information (i.e., blue = 1, red = 0), Condition 3 was environmental stability (i.e., whether the probability of the blue card being correct was stable (1), or volatile (0)). Condition 4 was social stability (i.e., whether the probability of the social information being correct was stable (1), or volatile (0)). Condition 5 was the interaction between reward value and environmental stability (i.e., Condition 1 x Condition 3), and Condition 6 was the interaction between social information and social stability (i.e., Condition 2 x Condition 4).

Final questions

Participants were also asked two yes or no questions related to the TD contagion task at the end of the experiment: *"Did you believe that the other player was a real participant?"*, and *"Did you notice yourself changing your own choices throughout this task?"*. These questions follow the same format as presented in previous research (Thomas et al., 2021).

Results

All analyses were run using Jamovi (version 1.6.15). Bayes factors (calculated using Jamovi jsq package presets) are reported for all analyses and are interpreted in line with the jsq package criteria. For all analyses, BF_{10} is used for significant analyses (measured by standard *p* values) and denotes the likelihood that the hypothesis of interest is correct. BF_{01} is used for non-significant analyses and denotes the likelihood that the null hypothesis is correct.

Social contagion in the temporal discounting task

To determine whether there was a significant change in behaviour after learning the value preferences of another person (i.e. a significant contagion effect, measured by normalised D_{KL}), a Wilcoxon signed-rank test was run (as data were not normally distributed; W = .93, p < .001). In line with previous research (Thomas et al., 2021), a significant contagion effect was found, and normalised D_{KL} was significantly greater than zero (t (187) = 12735.00, p < .001, d = .39, BF_{10}= 32931.88, **Figure 3**), with extreme Bayesian support for the finding. A correlation analysis revealed that participant's accuracy on this task (i.e., the number of trials on which the participant chose the same option as the modelled other) was not related to strength of contagion (r = -.02, p = .842, BF_{01}= 10.74).

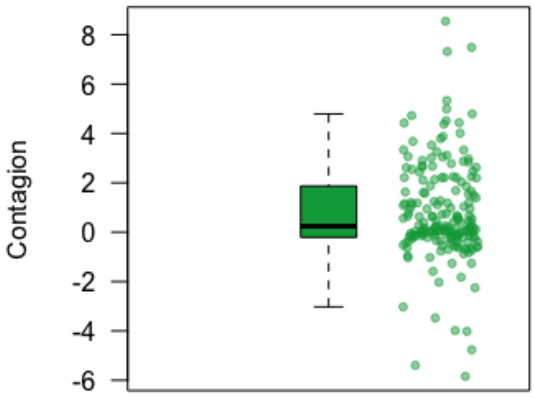


Figure 3. Contagion values. This shows descriptive statistics for normalised D_{KL} (i.e., contagion in the TD task).

In this task, 50% of participants believed that they were making choices on behalf of another real participant, and 70.75% of participants noticed their behaviour changing throughout the task. To determine whether belief in the manipulation, and awareness of behavioural change were related to contagion, two robust independent-measures t-tests were run on normalised D_{KL}. Belief in manipulation did not impact contagion (t (1112.99 = .67, p = .505, ξ = .07, BF_{01}= 6.23), and there was also no significant effect of awareness

of behavioural change on normalised D_{KL} (t (70.43) = .53, p = .602, ξ = .07, BF_{01} = 5.941e-9). Descriptive statistics for task parameters and variables related to the TD contagion task are presented in **Table 1**.

Table 1. Descriptive statistics for TD contagion task variables.

Variable	Statistic	
	M	*SD*
Log *k*	-1.99	.83
Log β	-.94	.04
D_{KL}	.78	2.01
Accuracy (out of 50)	39.21	5.25

Advice-taking in the card game

To determine whether participants were more likely to follow the advice (i.e., the previously chosen card) in the card game when the advice started out as helpful (i.e., congruent with the rewarded card), and whether there was a difference in advice-taking dependent on task stage, a 2 x 4 mixed-measures factorial ANOVA was run. Descriptive statistics for the advice-taking variables are presented in **Figure 4**.

There was a significant main effect of condition (i.e., helpful versus misleading; F (1, 186) = 28.16, p < .001, η^2 = .13, BF_{10} = 37008.45), with participants being more likely to follow the social advice across all stages if the advice started out as being helpful. There was also a significant main effect of stage (i.e., percentage of advice taking across all trials, volatile blocks, stable helpful blocks, and stable misleading blocks; F (3, 558) = 138.18, p < . 001, η^2 = .43, BF_{10} = 3.654e+62).

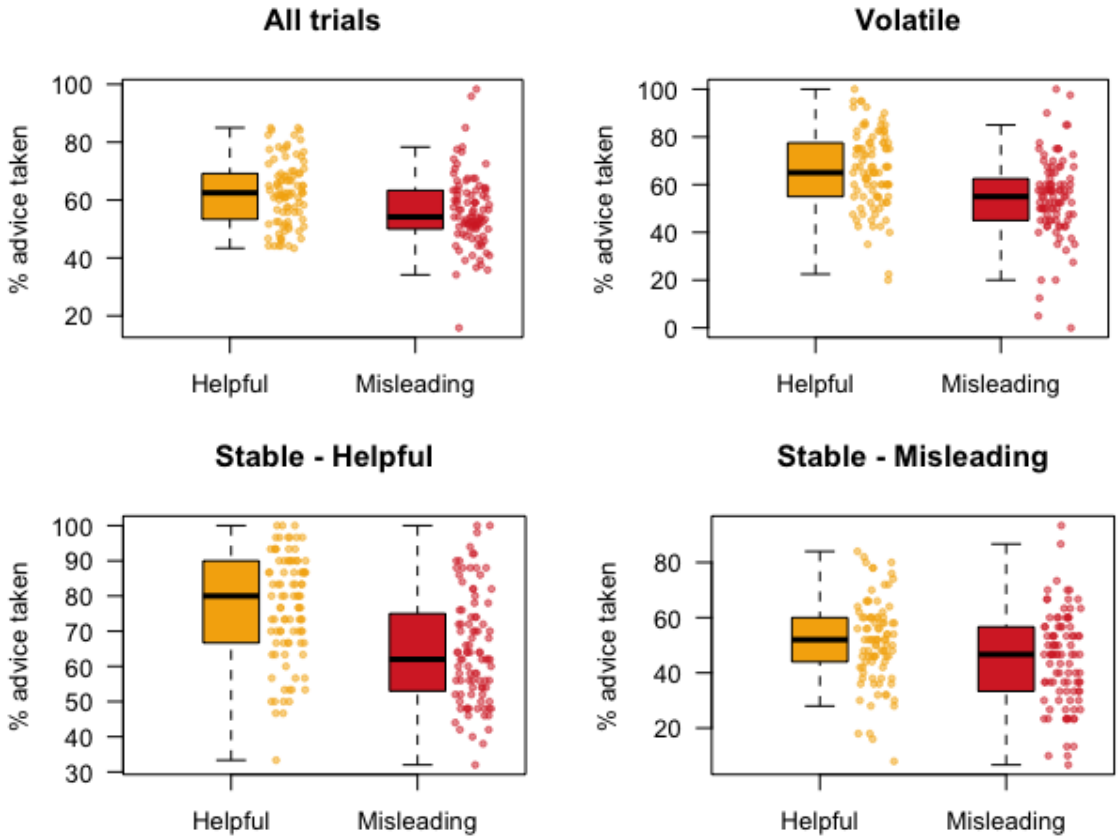


Figure 4. Percentage of advice taken in the card game task. This shows the percentage of trials on which the participant took the social advice across all trials, when the social information was volatile (trials 31-70), when the advice was stable and helpful (i.e., trials 1-30 for the helpful advice condition, and trials 71-120 for the misleading advice condition), and when the advice was stable and misleading (i.e., trials 71-120 for the helpful advice condition and trials 1-30 for the misleading advice condition, split by condition (i.e., helpful or misleading advice first).

Bonferroni post-hoc tests revealed that the difference between percentage of advice taken across stable helpful stages versus stable misleading stages was significant (t (558) = 20.34, $p < .001$, BF_{10} = 7.926e+29), and participants followed the advice in the stable and helpful stage more often than in the stable and misleading stage. This indicates that participants were able to successfully identify stages in which the advice was helpful or misleading and make use of the advice accordingly. There was also a significant difference in percentage of advice followed across the volatile and stable and

helpful stages (t (558) = -9.39, p < .001, BF_{10} = 2.255e+12), with participants following advice more in the stable and helpful stage. Participants were also more likely to follow advice in the volatile stage versus the stable and misleading stage (t (588) = 10.95, p < .001, BF_{10} = 8.771e+13). Together, these findings indicate that participants were also able to change their advice-taking behaviour to differentiate between stable and volatile blocks. There was no significant difference in advice-taking across all trials versus the volatile stage (t (588) = -.84, p = 1.00, BF_{01} = 5.75).

The relationship between advice-taking, contagion, and accuracy

To determine whether advice-taking on the card game was associated with contagion in the TD task, a hierarchical regression analysis was conducted. There were differences in advice-taking dependent on condition, so condition (i.e., helpful versus misleading advice first in the probability task) and task order (i.e., probability task versus TD task first) were entered as variables in the first block. Normalised D_{KL} was entered as the outcome variable, and the predictors in the second block were the four advice-taking variables (i.e., advice taken across all trials, advice taken in volatile trials, advice taken in stable and helpful trials, and advice taken in stable and misleading trials). Individual predictor statistics for all predictors are shown in **Table 2.**

The first block explained .01% of the variance in normalised D_{KL}, and was not a significant predictor (F (2, 185) = 2.21, p = .113, BF_{01} = 3.59). The second block also explained .01% of the variance in normalised D_{KL}, and was not a significant predictor overall (F (6, 181) = 1.32, p = .252, BF_{01} = 124.83). None of the individual advice-taking predictors significantly predicted D_{KL}, suggesting no relationship between advice taking and the contagion effect. Condition was a significant predictor independently with higher

contagion for participants who saw the helpful advice first versus those who saw the misleading advice first (see **Figure 5**), although a Bayesian independent-samples t-test indicated that support for an effect of condition on contagion was anecdotal (BF_{01} = 1.27). As there was not an effect of order of task, this finding is likely a chance finding.

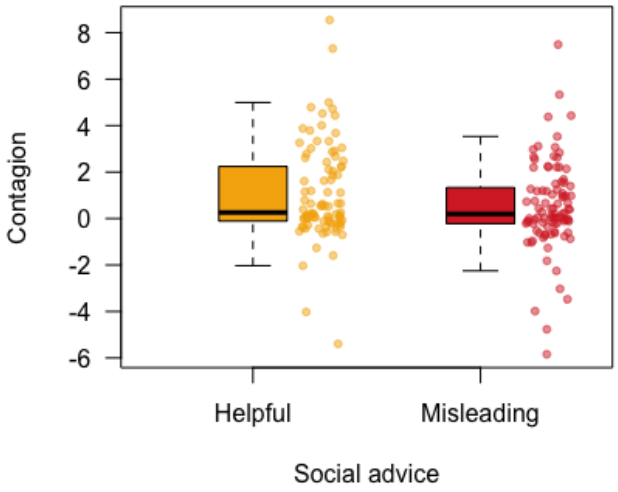


Figure 5. Contagion split by probability task condition. This shows contagion (measured by normalised D_{KL}) for participants in the helpful and misleading advice conditions on the probability task (i.e., helpful advice first, or misleading advice first).

A further hierarchical regression analysis was conducted to determine whether advice-taking in the card game task was associated with task accuracy on the TD task. Condition (i.e., did the probability task start with helpful or misleading advice for the first 30 trials) and task order were entered into the first block, and the advice-taking variables (i.e., all trials, volatile trials, helpful and stable, helpful and misleading) were entered as predictors in the second block. Individual predictor statistics for all predictors are shown in **Table 2**.

The first block explained .01% of the variance in accuracy, and was not a significant predictor of TD contagion task accuracy (F (2, 185) = 2.40, p = .093). The second block explained -.00% of the variance in accuracy and was not a significant predictor of TD contagion task accuracy overall (F (6, 181) = .90, p = .499). None of the individual predictors significantly predicted accuracy at making choices on behalf of the

other agent in the TD task, indicating that learning about the behaviour of the other in order to follow their behaviour was not related across tasks.

Table 2. Statistics for the individual predictors (condition, and the four advice-taking variables) entered into the regression model to predict normalised D_{KL} and accuracy at making choices on behalf of the other in the TD task.

Predictor	Normalised D_{KL}			Accuracy		
	β	t	p	β	t	p
Condition (helpful/misleading)	.87	2.30	**.022***	.90	.91	.363
Task order	.27	.92	.361	1.31	1.71	.089
Advice taken						
All trials	.04	.73	.466	-.02	-.12	.903
Volatile	-.02	-1.03	.304	.02	.30	.768
Stable helpful	-.02	-.71	.478	.01	.13	.898
Stable misleading	-.03	-1.07	.288	-.01	-.23	.821

Note: * is for $p < .05$.

The relationship between card game choice behaviour and contagion

To determine whether contagion was related to individual differences in environmental and social learning, correlation analyses were conducted on beta values derived from each participant. These values describe the extent to which reward utility, environmental and social stability, and the social information predicted participant choice on the probabilistic task (see the **Social probabilistic reward task variables** section). No significant relationships were observed between any of these beta values, contagion or accuracy. See **Table 3** for correlation statistics for these analyses.

Table 3. Correlation statistics for the relationships between beta values for each participant (produced from the regression analysis outlined in the **Social probabilistic reward task variables** section), and contagion (measured by normalised D_{KL}), and accuracy (out of 50) at making choices on behalf of the other in the TD task.

Variable	Normalised D_{KL}			Accuracy		
	r	p	BF_{01}	r	p	BF_{01}
1) Reward utility	.03	.676	10.05	.08	.255	5.76
2) Social choice	-.05	.538	9.08	.03	.646	9.87
3) Environmental stability	-.00	.980	10.95	-.00	.951	10.93
4) Social stability	-.01	.888	10.85	.07	.375	7.41
5) Reward * Stability interaction	-.04	.629	9.76	.08	.306	6.52
6) Social * Stability interaction	.03	.706	10.21	-.03	.720	10.28

HADS, and association with advice taking, contagion, and accuracy

Finally, exploratory correlation analyses were run to determine the effects of anxiety (HADS-A mean = 8.64, SD $\pm$ 3.93) and depression (HADS-D mean = 6.09, SD $\pm$ 3.54) on advice-taking in the card game, and accuracy and contagion in the TD task (critical *p* value adjustment would be made to allow for multiple comparisons in the event of a significant result). None of the relationships between anxiety and depression and task variables were significant, and the results of these analyses are displayed in **Table 4**.

Table 4. Correlation matrix with correlations presented between the anxiety and depression subscales of the HADS, and normalised D_{KL}, and accuracy in the TD task, and advice-taking in the probability task.

Task variable	HADS-A			HADS-D		
	r	p	BF_{01}	r	p	BF_{01}
TD accuracy	.08	.293	6.34	.03	.699	10.17
Normalised D_{KL}	-.04	.588	9.47	.01	.913	10.89
Advice taken						
All trials	.04	.630	9.76	-.02	.776	10.91
Volatile	-.00	.990	10.95	.04	.813	9.28
Stable helpful	.04	.590	9.49	.04	.748	9.23
Stable misleading	.01	.940	10.92	-.04	.190	9.66

Discussion

In this study, Bayesian modelling was used to investigate contagion of TD value preferences in an online sample of young adults. The relationship between contagion and choice behaviour on a probabilistic social reward learning task presented in the format of a card game was also explored. A significant effect of contagion was found, and participant's own TD preferences (as assessed by log *k*) became more similar to the preferences of the other agent. Neither contagion nor task accuracy (i.e., participants' accuracy at making choices on behalf of the modelled other) were associated with advice taking on the card game task. The extent to which participants choices were predicted by the social information, the reward probability, and environmental volatility was also not significantly related to contagion. This is despite both tasks being underpinned by Bayesian processes, suggesting that social learning behaviour in other tasks may not be associated with contagion in the TD task.

Contagion in an online sample

In line with previous research, we found a significant contagion effect, and learning the alternative value preferences of another agent produced significant changes in participant's own discount rates (Apps & Ramnani, 2017; Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Thomas et al., 2021). Whilst previous research was lab-based, our study was conducted online. Finding an effect of contagion in both lab-based and online studies indicates that contagion is a reliable effect which is reproducible across different contexts.

As only 50% of participants believed that they were viewing the choices of another real participant, this finding also supports previous research that found no differences in contagion, or the neural basis of contagion dependent on the identity of the other (Garvert et al., 2015; Thomas et al., 2021). Research also shows that individuals are influenced by the norms of an overall social group, as well as the behaviour of individual human or computer agents (e.g., Apps & Ramnani, 2017; Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Park et al., 2017; Thomas et al., 2021; Suzuki et al., 2016), indicating that contagion of value preferences is a strong and reproducible effect, regardless of the identity of the other agent.

Contagion and learning

Research has now repeatedly shown that contagion occurs regardless of whether participants think that the other agent is real or not (Garvert et al., 2015; Thomas et al., 2021). This raises the question as to what extent contagion is a social effect, or whether it is simply an informational effect of learning about another agent. Our previous study found no relationship between contagion and scores on the Autism Quotient (AQ;

Thomas et al., 2021), although it is possible that questionnaire measures are not suited to exploring individual differences in the contagion effect. Previous research has indicated that it is difficult to link individual differences in questionnaire measures with task data due to differences in variability, with lower variability in behavioural tasks versus questionnaires (Frey et al., 2017; Hedge et al., 2017; Palminteri & Chevallier, 2018; Pedroni et al., 2017). Here, we endeavoured to establish whether contagion was associated with either social or environmental learning using the social probabilistic reward learning paradigm. Using this task, it was possible to extract parameters about how quickly participants learned the environmental properties (i.e., which card pays out) as well as the social learning (i.e., is the advice good or bad?) and see if either of these predicted individual differences in contagion. We found no relationships between either usage of the social or non-social information (i.e., social and non-social reward probability and stability/volatility), indicating no relationship between learning and contagion across tasks. However, future research could also focus on using more in-depth computational modelling (as per Behrens et al., 2007; 2008; Henco et al., 2020 and Sevgi et al., 2020) to calculate learning rates and determine how successfully participants were able to learn about the probabilities associated with social and non-social information, and volatility. This research could also determine whether learning rates are associated with contagion in the TD task.

We also did not find an association between contagion in the TD task, and advice-taking in the probabilistic social learning task, providing further support for a lack of a relationship between tasks. We did find that participants were more likely to follow advice when the advice started out as helpful, indicating that participants did learn about whether the advice was helpful or not. In further support of this, there were also

differences in advice-taking between stages, such that participants took the advice more when it was helpful, versus when it was volatile, or misleading. Despite participants exhibiting this ability to differentiate between stages (and therefore successfully learn about the behaviour of the other agent), we did not find an association between advice-taking at any stage of the task, and contagion. We also did not find an association between accuracy and contagion in the TD task. Whilst previous research has been found that contagion is associated with accuracy (Reiter et al., 2019), and is driven by learning signals such as prediction errors (PEs; Garvert et al., 2015; Suzuki et al., 2016), these findings suggest that the relationship between behavioural measures of learning and contagion may not be reliable.

However, as contagion in this TD task is such a strong effect (with extreme Bayesian support for contagion now found here and in previous research by Thomas et al., 2021), it is possible that differences in other behavioural measures of learning have little impact on this effect. Contagion and social influence effects such as conformity do appear to be driven by learning signals such as PEs, as well as weighting of social information, and multiple tasks have now found differences in these effects across participants (e.g., Apps & Ramnani, 2017; Garvert et al., 2015; Park et al., 2017; Suzuki et al., 2016). As a reliable association between learning and contagion has not been found, it would be of interest to determine whether learning related activation, such as PE signalling, or activation related to differential weighting of social and non-social information (e.g., participant's own beliefs, as well as reward-related signalling), is associated across tasks assessing contagion and social learning.

Conclusion

In summary, these results indicate that contagion is a reliable and reproducible effect, which extends to behavioural studies in laboratory and online settings. These results also indicate that contagion may be an effect that it is independent of propensity for learning on other behavioural tasks, although further research should aim to determine whether this behavioural change and learning is related to neurobiological indicators of behavioural change across tasks.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*(4), 463-496.

Apps, M. A. J., & Ramnani, N. (2017). Contributions of the medial prefrontal cortex to social influence in economic decision-making. *Cerebral Cortex, 29*(9), 4635-4648.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1-70.

Basile, A. G., & Toplak, M. E. (2015). Four converging measures of temporal discounting and their relationships with intelligence, executive functions, thinking dispositions, and behavioral outcomes. *Frontiers in Psychology, 6*, 728.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*(7219), 245-249.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214–1221.

Beltzer, M. L., Adams, S., Beling, P. A., & Teachman, B. A. (2019). Social anxiety and dynamic social reinforcement learning in a volatile environment. *Clinical Psychological Sciences, 7*(6), 1372–1388.

Bennett, D. (2015). The neural mechanisms of Bayesian belief updating. *The Journal of Neuroscience, 35*(50), 16300-16302.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology, 20*(13), 1165-1170.

Crawford, J. R., Henry, J. D., Crombie, C., & Taylor, E. P. (2001). Brief report: Normative data for the HADS from a large non-clinical sample. *British Journal of Clinical Psychology, 40*, 429–434.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgements according to its reliability. *The Journal of Neuroscience, 37*(25), 6066-6074.

Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J. & Dolan, R. J. (2015). Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron, 85*(2), 418-428.

Henco, L., Diaconescu, A. O., Lahnakoski, J. M., Brandi, M. L., Hörmann, S., Hennings, J., … Mathys, C. (2020). Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder. *PLoS Computational Biology*, *16*(9), e1008162.

Karvelis, P., Seitz, A. R., Lawrie, S. M. & Seriè, O. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife, 7,* 1-29.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86.

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience & Biobehavioral Reviews, 69*, 313-332.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience, 20*(9), 1293-1299.

McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J. C., & Stout, S. C. (2009). A comparison of four models of delay discounting in humans. *Behavioural Processes, 81*(2), 256–9.

Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of inter-temporal preferences. *PLoS Computational Biology, 12*(7), 1-17.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114-1121.

Park, S. A., Goïame, S., O'Connor, D. A., & Dreher, J. C. (2017). Integration of individual and social information for decision-making in groups of different sizes. *PLoS Biology, 15*(6), 3-28.

Perreault, C., Moya, C., & Boyd, R. (2012). A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior, 33*, 449-459.

Person, B., Sy, F., Holton, K., Govert, B., Liang, A., & National Center for Infectious Diseases/SARS Community Outreach Team (2004). Fear and stigma: The epidemic within the SARS outbreak. *Emerging infectious diseases, 10*(2), 358–363.

Rajkumar, R. P. (2020). COVID-19 and mental health: A review of the existing literature. *Asian Journal of Psychiatry, 52*, 102066.

Reiter, A. M. F., Suzuki, S., O'Doherty, J. P., Li, S. C. & Eppinger, B. (2019). Risk contagion by peers affects learning and decision-making in adolescents. *Journal of Experimental Psychology: General, 148*(9), 1494-1504.

Shultz, J. M., Cooper, J. L., Baingana, F., Oquendo, M. A., Espinel, Z., Althouse, B. M., … Rechkemmer, A. (2016). The role of fear-related behaviors in the 2013-2016 West Africa Ebola virus disease outbreak. *Current Psychiatry Reports, 18*(104), 1-14.

Sevgi, M., Diaconescu, A. O., Henco, L., Tittgemeyer, M., & Schilbach, L. (2020). Social Bayes: Using Bayesian modeling to study autistic trait–related differences in social cognition. *Biological Psychiatry, 87*(2), 185-193.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., … Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron, 74*, 1125-1137.

Suzuki, S., Jensen, E. L. S., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioural contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *PNAS, 113*(14), 3755-3760.

Thomas, L., Lockwood, P. L., Garvert, M. M., & Balsters, J. H. (2021). Contagion of temporal discounting value preferences in neurotypical and autistic adults. *Journal of Autism and Developmental Disorders.* Advance online publication.

Vilares, I., & Kording, K. (2011) Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences, 1224*(1), 22–39.

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the

2019 Coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environmental Research and Public Health, 17*(5), 1729.

Wheeler, L. (1966). Toward a theory of behavioral contagion. *Psychological Review, 73*(2), 179-192.

Whitton, A. E., Treadway, M. T., & Pizzagalli, D. A. (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry, 28*(1), 7–12.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavia, 67*(6), 361-370.

Chapter 7.

General Discussion

7.1. Thesis overview

Previous research has shown that our own subjective valuations can be altered through contagion (Burke et al., 2010; Campbell-Meiklejohn et al., 2010; Garvert et al., 2015; Klucharev et al., 2009; Nicolle et al., 2012; Suzuki et al., 2016; Reiter et al., 2019; Zaki et al., 2011). The overarching aim of this PhD thesis was to investigate individual differences in contagion of value preferences in TD and PD tasks. The four experimental chapters presented within this thesis each use a Bayesian method of exploring contagion, and each explore different aspects of individual differences in the effect.

7.2. Chapter summaries – Main research questions

7.2.1. Contagion of temporal discounting value preferences in neurotypical and autistic adults

In this chapter, three studies were conducted in order to explore individual differences in contagion of TD preferences in two independent NT samples, and one sample of autistic adults. The TD contagion task used in all three studies was composed of five blocks of trials, in which participants made choices on behalf of themselves, and two modelled agents (i.e., Self1, Other1, Self2, Other2, Self3). The behaviour of these agents was modelled to be either more impulsive than the participant (i.e., with a discounting parameter, log *k*, that was +1 that of the participant), or more patient (i.e., with a log *k* that was -1 that of the participant), although participants were informed that these were the choices of previous participants. In all three studies, participants completed the TD contagion task, as well as a measure of autistic traits. A matched subset of NT participants from the Study 2 sample were also selected for comparison with the Study 3 sample.

The key research questions for this chapter were as follows:

(1) Do NT individuals shift their preferences when they learn about other people's preferences (i.e., contagion)?

(2) Are levels of autistic traits in the NT population associated with contagion of value preferences?

(3) Do autistic adults show comparable shifts in preference to NT controls?

7.2.2. Does contagion generalise across different tasks? A comparison of temporal and probability discounting

In this chapter, the differences and similarities in contagion and task parameters between TD and PD tasks were explored. A sample of NT adults (the same sample as in **chapter 3**, Study 2) completed a measure of autistic traits, as well TD and PD contagion tasks. Each discounting task had five blocks of trials (i.e., Self1, Other1, Self2, Other2, Self3), and participants made choices for themselves, and on behalf of two modelled others (i.e., a more impulsive/positive and a more patient/negative other in the TD task, and a more risk-aversive/positive and a more risk-seeking/negative agent in the PD task).

The key research questions addressed in this chapter were as follows:

(1) Do NT individuals shift their preferences when they learn about other people's preferences (i.e., contagion) in both TD and PD tasks?

(2) Is the magnitude of contagion similar across tasks?

(3) Are discounting task parameters related across TD and PD tasks?

(4) Are levels of autistic traits in the NT population associated with contagion of value preferences in TD and PD tasks?

7.2.3. The power of friendship: Does social distance modulate neural pathways controlling social contagion and learning?

In this chapter, the effect of social distance on contagion was explored. Pairs of adult participants (i.e., pairs of close friends or partners) that had known each other for six months or more, made TD choices on behalf of themselves, each other, and a stranger (i.e., a confederate). Following a short training task, in which participants completed choices only for themselves, the participants each made choices on behalf of themselves and the two other agents (i.e., Close and Distant Others) whilst in the scanner. Participants believed that the choices from the training session were used in the scanner, but these choices were modelled. The scanner task had five blocks of trials: Self1, Other1, Self2, Other2, Self3. On each trial, the options were first displayed as an Offer screen, and participants could think about which option they preferred at this stage. The options were then displayed for the Choice stage, and participants made their selection here. On all Other-trials participants also then saw a Feedback screen, which informed them whether they had made the same choice as the modelled other. Both whole brain and ROI analyses were conducted (with the ROIs specified according to the areas identified by Garvert et al., 2015). ROI related to contagion were in the mPFC and STS, and ROIs related to feedback processing were in the pMFC, insula, and striatum.

The key research questions addressed in this chapter were as follows:

(1) Do participants shift their own TD value preferences when they learn about other people's preferences (i.e., contagion)?

(2) Is contagion stronger for close others versus distant others?

(3) Can a brain-based effect of contagion be observed?

(4) Are there differences in the BOLD response for contagion and learning for close and distant others?

7.2.4. Temporal discounting contagion and probabilistic social learning: Is contagion linked to advice-taking?

In this chapter, an online study was conducted in order to determine whether contagion of TD preferences was associated with choice behaviour and advice-taking in a probabilistic social reward learning task, presented as a card game. Participants completed a TD contagion task, with three blocks (i.e., Self1, Other, Self2), and the behaviour of the other agent (i.e., more impulsive or more patient than the participant) was counterbalanced across participants. The card game included additional social information, and both the reward and social information went through volatile and stable phases, with half of the participants being shown social information that was helpful at the start of the experiment (i.e., social information that showed the rewarded card), and the other half being shown social information that was misleading at the start of the experiment. Participants also completed measures of anxiety and depression.

The key research questions for this chapter were as follows:

(1) Do participants shift their own TD value preferences when they learn about other people's preferences (i.e., contagion)?

(2) Is contagion associated with advice-taking and choice behaviour on the card game?

(3) Are contagion, accuracy at making choices on behalf of the other in the TD task, and choice behaviour on the card game associated with self-reported anxiety and depression?

7.3. Discussion of findings

7.3.1. Contagion effects on discounting preferences

Despite previous research suggesting that TD and PD preferences are stable, and can be considered as stable personality traits across time periods as long as a year (Kirby, 2009; Ohmura et al., 2006), significant contagion effects were observed in all four experimental chapters, indicating that contagion is a strong and reproducible effect. See **Figures 1**, **2**, **3**, **4**, which are included throughout this section to allow for direct comparison across chapters.

In the behavioural studies presented in **chapter 3**, a strongly significant contagion effect (as measured by D_{KL}) was found for the more patient (i.e., negative) agent on the TD task in both NT samples (i.e., Study 1 and Study 2 samples). For the more impulsive (i.e., positive) other agent, a significant contagion effect was found only in the Study 1 NT sample, although Bayesian support for this effect was anecdotal. The difference in contagion for patient versus impulsive others was only significant in the Study 1 sample, although the Bayes factors for both analyses provided only anecdotal support for the findings. Findings from the sample of autistic participants included in this chapter (i.e., Study 3 sample) are outlined and discussed in **section 7.3.5.** See **Figure 1** for descriptive statistics for the contagion effects in this chapter.

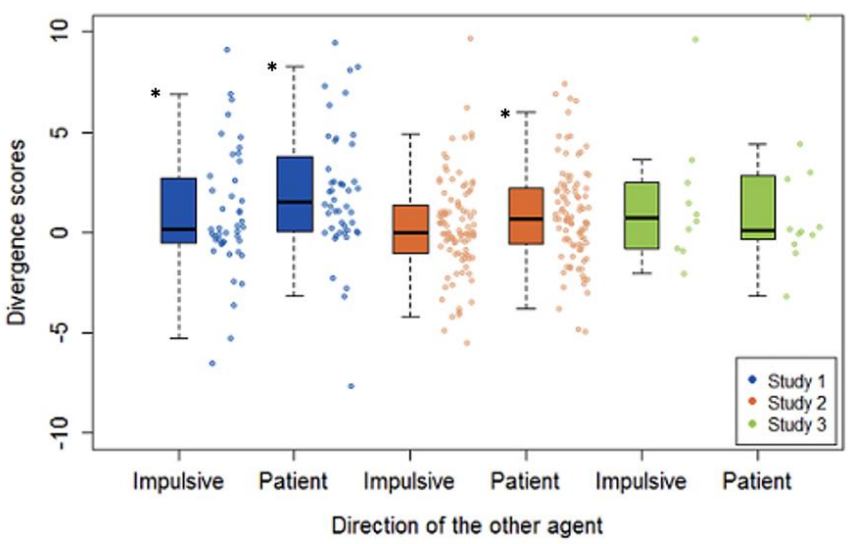


Figure 1. Descriptive statistics for chapter 3 D_{KL}. This shows the contagion effects in the three study samples (i.e., Study 1 NT, Study 2 NT, and Study 3 ASC) included in **chapter 3**. This figure is also presented as **Chapter 3 Figure 3.**

In **chapter 4**, the finding of a significant contagion effect for increasingly negative but not increasingly positive agents was also found behaviourally in the PD contagion task, with a significant contagion effect found for risk-seeking (i.e., negative), but not risk-aversive (i.e., positive) agents. The effect of direction was also not significant in **chapter 4**, with no significant difference in contagion for negative (i.e., patient/risk-seeking) versus positive (i.e., impulsive/risk-aversive) agents. This bias towards becoming more patient replicates the behavioural findings of Moutoussis and colleagues (2016), who were unable to attribute this bias to differences in any model parameters on the TD task. Research into PD has found contagion for both risk-seeking and risk-aversive agents, with the difference between these two not reaching significance (Suzuki et al., 2016), which is not supported by the findings of **chapter 4**. As discussed in **chapter 4**, the effect observed here could be due to the preferences of more impulsive and more risk-aversive agents appearing more extreme to participants (given that the majority of adults tend towards being risk-averse (Paulsen et al., 2012), and the participant behaviour on the TD task being closer to impulsive than patient). Suzuki and colleagues (2016) previously modelled the behaviour of other agents with fixed choices,

meaning that the difference between participants' own preferences and the preferences of the agent would have varied across participants. Whilst holding this constant reduces individual differences, it is possible that this variation meant that the behaviour change was less noticeable for participants than in **chapter 4**. Future research could determine whether varying the extent to which the behaviour of the other agent differs from the behaviour of participants impacts upon contagion. For example, future research could determine whether a contagion effect is found for both more negative (i.e., patient/risk-seeking) and more positive (i.e., impulsive/risk-aversive) agents when the other agent has a log k/α that is ± 0.5 that of the participant, instead of ± 1. See **Figure 2** for descriptive statistics for the contagion effects in this chapter.

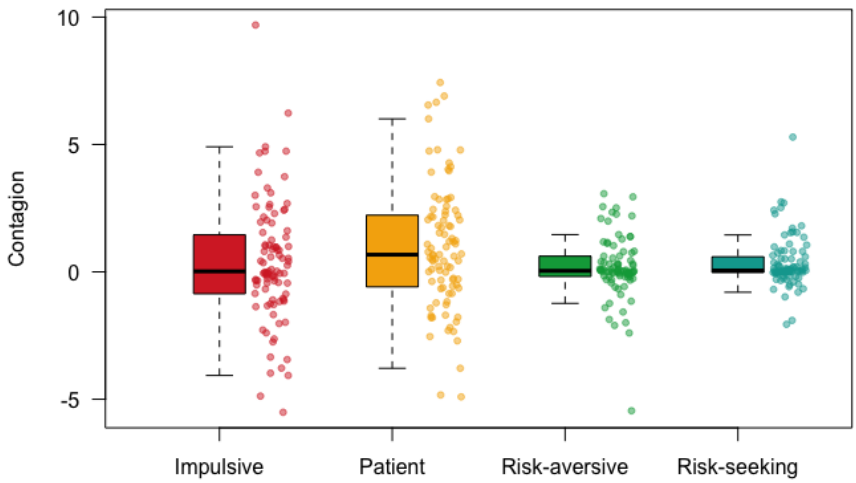


Figure 2. Descriptive statistics for chapter 4 D_{KL}. This shows the contagion effects for TD and PD tasks in **chapter 4**. This figure is also presented as **Chapter 4 Figure 3.**

Throughout the thesis, a Bayesian method of indexing contagion was used (as per Suzuki et al., 2016), as previous research indicates that Bayesian models of social influence (which account for the participant's degree of certainty in their initial belief, as well as the credibility of the newly presented social information) outperform other models of social influence (De Martino et al., 2017; Moutoussis et al., 2016; Park et al., 2017; Suzuki et al., 2016). The method used to measure contagion within this thesis is Kullback-Leibler divergence (D_{KL}), which accounts for both the overall change in

participant's preferences, as well as their degree of certainty in their initial belief, and their degree of certainty in the posterior belief, after making choices on behalf of the other (Kullback & Leibler, 1951).

In **chapter 3** and **4**, D_{KL} was compared with contagion measured by the same shift variable used by Garvert and colleagues (2015) to measure contagion (i.e., the distance between the discounting parameter values at the end of the two self-blocks, divided by the distance between the discounting parameter of the modelled other, and discounting parameter at the end of the first self-block). Whilst this measure does account for the change in value preference controlled by the distance that participants were able to change, this variable does not account for participants' uncertainty in their belief in both blocks. In **chapter 3**, the bias towards becoming more patient on the TD task was replicated in both NT samples when using the shift variable, and in **chapter 4**, the same bias towards becoming more risk-seeking on the PD task was also found. In both **chapter 3** and **4**, contagion effects were stronger (with stronger Bayesian support) in the majority of conditions when using the D_{KL} variable instead of the distance-controlled shift variable, indicating that accounting for uncertainty produced a stronger effect. In future research, individual differences between uncertainty (and the extent to which this drives contagion) could be explored in both TD and PD tasks.

Overall, participants choices in the first Self-block were noisier in the PD task than in the TD task (as indicated by β values, with extreme Bayesian support for the effect). As choices were noisier in this task, this may indicate that participants were less certain about their own PD choices than their TD choices. Although Bayesian support for an effect of contagion was lower in the PD task (with smaller Bayes factors) which is at odds with previous findings indicating that individuals are more likely to change

their beliefs when they are uncertain about these (De Martino et al., 2017; Moutoussis et al., 2016; Park et al., 2017). It is possible that uncertainty affects TD and PD preferences differently, providing further reasoning for exploring the extent to which uncertainty impacts upon contagion differs across TD and PD tasks in future research.

In **chapter 4** contagion for more patient agents in the TD task and more risk-seeking agents in the PD task was significantly positively correlated, indicating that participants' propensity to make choices that aligned with more negative discounting parameter values was related across tasks. Previous neuroimaging studies have found similarities and differences in discounting across tasks, with the OFC and ventral striatum tracking stimulus value in both tasks (Bartra et al., 2013; Levy & Glimcher, 2012; Peters & Büchel, 2009), regions in the dmPFC showing specific activation for TD only, and regions in the parietal lobe showing activation for PD (Peters & Büchel, 2009). Within this thesis, both similarities and differences were found behaviourally. Whilst there were differences in the amount of noise in participants' own choices, examining the discounting parameters on each task also revealed a correlation between participants' own log k and log α values in the first Self-block, before they made choices on behalf of the other agents. Participants that were more patient on the TD task, were also more-risking on the PD task, although the Bayesian support for this effect was anecdotal. This finding supports the findings of Hayden and Platt (2007) but does not support the findings of Ohmura and colleagues (2006), who did not find a relationship between discounting parameters across tasks. Whilst the Bayes factor supporting a correlation between discounting parameters was anecdotal, finding a relationship between discounting parameters, as well as a correlation with strong Bayesian support between contagion for

patient and risk-seeking others, suggests a commonality of contagion and discounting across domains.

In the online study presented in **chapter 6**, the order of the behaviour of the other agent was counterbalanced across participants, and a significant contagion effect was found, replicating the findings of the previous chapters (see **Figure 3** for descriptive statistics for the contagion effect in this chapter). Here, the relationship between contagion and anxiety and depression (as measured by the HADS; Zigmond & Snaith, 1983) was also explored, although no association was found between self-reported anxiety and depression, and propensity for contagion. In the fMRI study presented in **chapter 5**, contagion was split for close and distant others, and the variables included other agents with behaviour in both directions (as per Garvert et al., 2015). A significant contagion effect was found for both close and distant others, with no significant difference in contagion for the two agents (See **Figure 4** for descriptive statistics for the contagion effects in this chapter). Contagion was also statistically equivalent across close and distant others, indicating that the degree of contagion was similar regardless of identity in this sample (see **section 7.3.4.** for a discussion of the effect dependent on agent). In this chapter, both the ROI and whole brain analyses revealed no brain areas in which the BOLD signal was associated with contagion for close or distant others.

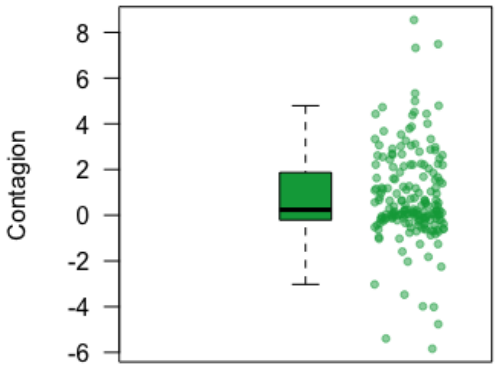


Figure 3. Descriptive statistics for chapter 6 D_{KL}. This shows the contagion effects for the TD task in **chapter 6** (which includes contagion for both more impulsive and more patient agents). This figure is also presented as **Chapter 6 Figure 3.**

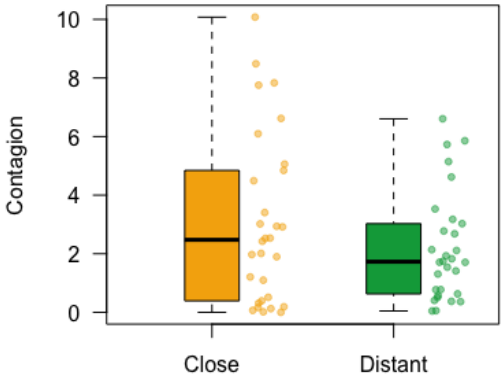


Figure 4. Descriptive statistics for chapter 6 D_{KL}. This shows the contagion effects for the TD task in **chapter 5** (contagion for both Close and Distant Others includes contagion for both more impulsive and more patient agents). This figure is also presented as **Chapter 5 Figure 5.**

Garvert and colleagues (2015) used repetition suppression to index the contagion effect in a TD task, whereas Suzuki and colleagues (2016) found a contagion effect without using repetition suppression. This may highlight further differences between TD and PD, although future research exploring different methods of indexing contagion in brain imaging experiments is required. It is also possible that methods exploring functional and structural connectivity could relate to individual differences in contagion in discounting tasks. Although an effect of contagion of value preferences was not observed in the neuroimaging data in **chapter 5**, Garvert and colleagues (2015) had previously found that differences in striatal surprise signals predicted the extent of value update by influencing signals in the mPFC. The absence of a brain-based effect could reflect methodological issues. For example, previous research using ICA found associations between neural oscillations in various functional networks and working memory task performance when traditional GLM based methods failed to find any differences in brain activity (Balsters et al., 2013). Group differences in cognitive performance dependent on white matter connectivity pathways in the cingulum have also been found (Bathelt et al., 2019). It is possible that differences in contagion both within and between tasks may also elucidated using different methods. Future research could

seek to determine whether network connectivity and differences in signal noise within networks are associated with contagion.

7.3.2. Learning the value preferences of others

In all chapters, learning of the value preferences of the other agents was measured using accuracy (i.e., the number of trials on which the participant chose the option that was preferred by the modelled other). In **chapter 3**, accuracy was higher for more patient agents in both NT samples, although percentage accuracy was slightly higher in the Study 1 sample (with 50 trials per block) versus the Study 2 sample (with 30 trials per block), indicating that increasing the number of trials may also increase accuracy. Further research could vary the number of trials in each block to determine whether this impacts upon accuracy in discounting tasks. Percentage accuracy is plotted for all four experimental chapters in **Figure 5**, to allow for direct comparison across tasks.

Greater accuracy for more patient agents on the TD task and more risk-seeking agents on the PD task was found in **chapter 4**, indicating that NT participants were more accurate at making choices on behalf of other agents with increasingly negative discounting parameters on both tasks. Participants were also more accurate at making choices on behalf of the other agents in the PD task versus the TD task, although Bayesian support for this effect was anecdotal.

Accuracy was also calculated for close and distant others in **chapter 5**, and participants chose the option preferred by the modelled other more frequently for the close other, than the distant other, providing support for a social distance effect on learning the preferences of others. This social distance effect was also underpinned by differences in the BOLD response during the feedback stage. The ROI analysis revealed differences in activation related to surprise at the choices of the other (i.e., PE), with

greater activation for close versus distant others, and the whole brain analysis also revealed one cluster with significantly greater activation associated with feedback for close versus distant others in the postcentral gyrus. These findings indicate that (despite a lack of a social distance effect on contagion) participants were more accurate at choosing on behalf of people that they were close to, versus strangers, and this behavioural effect was underpinned by activation associated with feedback.

As discussed in **chapter 5**, greater activation in the postcentral gyrus may be related to greater activation in response to imagining the choices of close versus distant others (Lamm et al., 2007). Future research could test for this hypothesis by asking participants to provide ratings of how easily they could imagine each participant making the TD choices, and by explicitly asking them to imagine the agents executing these choices. Finding a social distance effect on PE signalling in the pMFC (i.e., related to surprise at the choice of the other agents) along with a behavioural social distance effect on accuracy also suggests that learning in the preferences of the other agents in the TD task was driven by PE signalling. This is supported by previous research showing that learning is driven by PE signals in multiple brain areas (e.g., Balsters et al., 2017; Behrens et al., 2007; 2008; Garvert et al., 2015; Lee & Harris, 2013; Reynolds & Wickens, 2002), and that these signals are affected by social distance (Carp et al., 2008; Fareri et al., 2012; Kang et al., 2010; Powers et al., 2016). The social distance effect on learning and feedback processing here may be attributable to greater credibility attributed to the choices of close versus distant others, in line with previous research suggesting that the credibility of social information drives social influence effects (e.g., De Martino et al., 2017; Park et al., 2017).

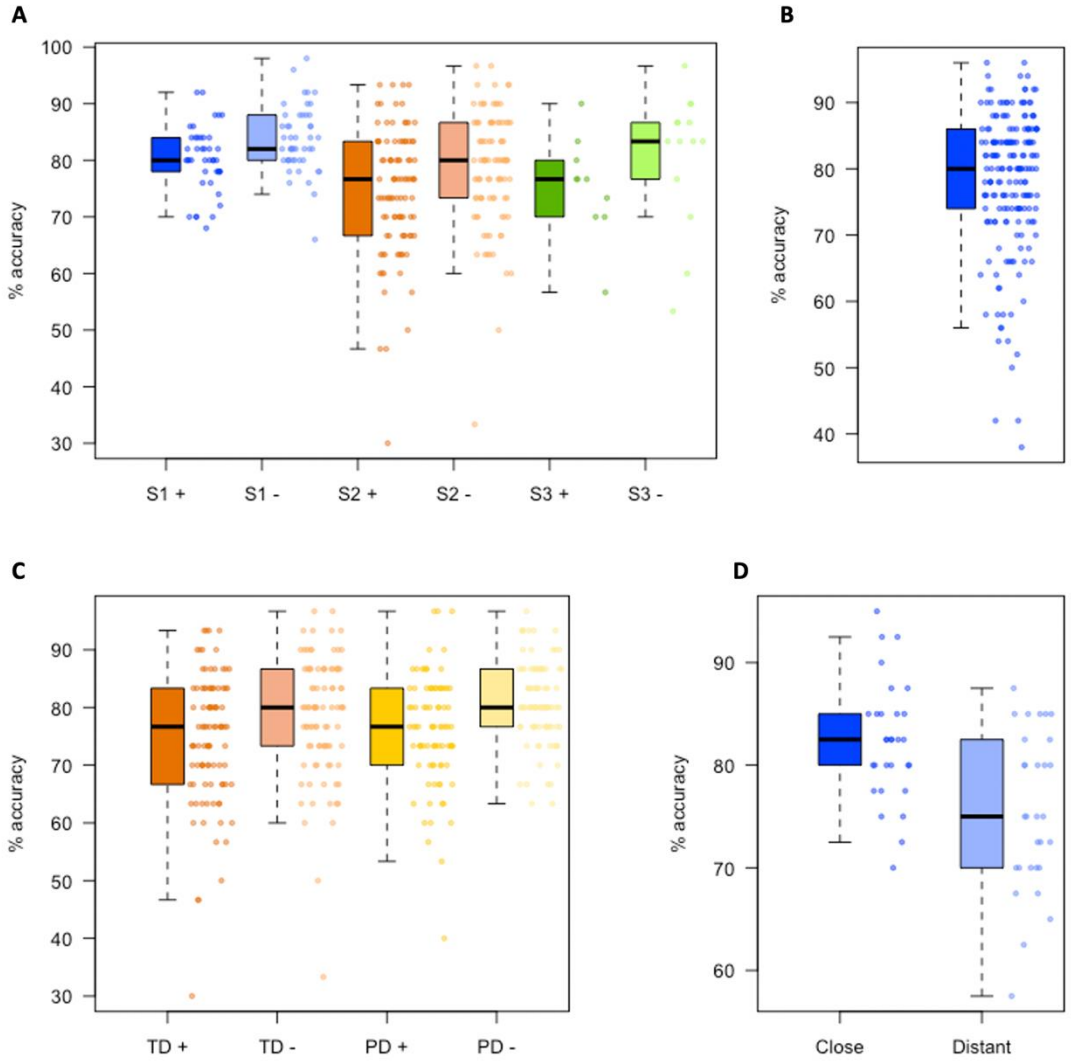


Figure 5. Percentage accuracy on for all four experimental chapters. A) Shows percentage accuracy for each Study sample (i.e., Study 1 NT, Study 2 NT, Study 3 ASC) and direction (i.e., positive/impulsive and negative/patient) in **chapter 3** (TD task). B) Shows percentage accuracy for the sample presented in **chapter 6** (TD task). C) Shows percentage accuracy for the TD and PD tasks in **chapter 4**, split by direction (i.e., positive: impulsive/risk-aversive and negative: patient/risk-seeking). D) Shows percentage accuracy in the TD task presented in **chapter 5**, split for Close and Distant Others.

In **chapter 6**, behavioural learning-related measures were also computed for the card game task. Overall, participants were more likely to follow social advice across all trial types when the social advice started out as being helpful (i.e., was congruent with the rewarded card) versus misleading, indicating that participants learned about how

useful the social information was at the beginning of the task. Participants were also more likely to follow advice in the stable and helpful phase of the task (i.e., when the advice was more likely to be correct) versus the stable and misleading and volatile stages of the task. Advice was also more likely to be followed in the volatile stage of the task versus the stable and unhelpful stage. Finding differences in advice-taking across blocks indicates that participants successfully tracked how useful the social information was at different points throughout the experiment, and made use of the advice accordingly. Future research could focus on using more in-depth computational modelling (as per Behrens et al., 2007; 2008; Henco et al., 2020 and Sevgi et al., 2020) to determine whether the extent to which participants use the social information and volatility guides choice behaviour, and how successfully participants were able to learn about the probability and volatility of the card and social information, as determined by learning rate.

7.3.3. The relationship between learning and contagion

Previous research has found that contagion is driven by the strength of PE signalling (Garvert et al., 2015; Suzuki et al., 2016). Some researchers have also found that behavioural contagion is associated with accuracy at making choices on behalf of others in a PD task (Reiter et al., 2019), although these researchers did not use a measure of contagion that accounted for uncertainty. The same relationship between contagion and accuracy was not observed in an earlier study that used the D_{KL} method of indexing contagion in a PD task (Suzuki et al., 2016).

In **chapter 3** of this thesis, only the correlation between accuracy and contagion (as measured by D_{KL}) for the more impulsive (i.e., positive) agent in the Study 2 sample was significant, and no significant correlations were observed between accuracy and

contagion for patient (i.e., negative) others in the Study 2 sample, or patient or impulsive others in the Study 1 sample. In **chapter 4**, no significant correlations were observed for risk-aversive (i.e., positive) or risk-seeking (i.e., negative) others in the PD task. In **chapters 5** and **6**, the finding of no relationship between behavioural measures of accuracy and contagion on the TD task was replicated. A significant relationship between the BOLD response associated with feedback and behavioural contagion was also not observed in **chapter 5**. Together, these findings suggest that there is not a reliable relationship between accuracy or feedback processing, and a measure of contagion which accounts for uncertainty in TD and PD tasks.

In **chapter 6**, measures of contagion and accuracy at making choices on behalf of the other agent in the TD task were not associated with participants' advice-taking behaviour in the card game in any stage of the task. This is despite participants differentiating between stages in the card game task (i.e., social information volatile, social information stable and helpful, and social information stable and misleading). The extent to which participants' choices were predicted by the social information, the reward probability, and volatility was also not significantly related to TD contagion. These findings suggest that although both tasks are underpinned by Bayesian associative processes, contagion may not be associated with behavioural indexes of learning on the card game task. In the previous section, I suggested that research could use computational modelling to determine whether learning rate varied across different stages of the experiment (as per Behrens et al., 2007; 2008; Henco et al., 2020 and Sevgi et al., 2020), and this research could also seek to determine whether learning rate is associated with contagion.

7.3.4. Is contagion a social effect?

Whilst significant contagion effects have been found throughout this thesis, replicating the findings of previous studies (Garvert et al., 2015; Moutoussis et al., 2016; Nicolle et al., 2012; Suzuki et al., 2016), previous research has also indicated that contagion in discounting tasks may a general informational effect of learning about another agent and is not necessarily a social effect (Ruff et al., 2013). Both Garvert and colleagues (2015) and Suzuki and colleagues (2016) found no differences in contagion when participants were informed that they were deciding on behalf of a human agent versus a computer agent. These findings are also supported by the findings of the experiments reported in this thesis.

In all four chapters, participants were asked whether they believed that they were making choices on behalf of real participants. This question was asked to determine whether participants were aware that the choices were produced by a computer agent (i.e., modelled), or had believed the experimental manipulation (i.e., had believed that the other players were human participants). In **chapter 3**, these questions were added to the Study 2 task, and 40.23% of participants that completed this measure believed that the other was a real participant. In **chapter 4**, 39.33% of participants believed the manipulation, and in **chapter 6**, 50% believed that they were making choices on behalf of a real participant. This was close to chance in all three studies, and belief in the manipulation did not affect TD or PD contagion in any of the studies presented in these chapters. This indicates that participants were influenced by the choices of the other to the same degree, regardless of whether they believed that the other participants were real, and suggests that contagion may not be a social effect.

However, participants in **chapters 3**, **4**, and **6** were not introduced to the other participants, which may have contributed to participants not believing that they were making choices on behalf of real participants. To further test for the existence of a social effect on contagion, participants in **chapter 5** made choices on behalf of real-world close and distant others. In this chapter, pairs of close friends or partners were recruited to take part in the experiment and were also introduced to strangers (i.e., experimental confederates). All three participants completed the training task together, before participants entered the scanner, and were told that they would be making choices on behalf of their testing partner and the stranger in their scanner, and should try to accurately guess their preferences (as indicated in the training task; although these choices were still modelled). Using real-world pairs of close and distant others successfully increased participants' belief in the experimental manipulation, and 90% of participants in this study believed that they were viewing the choices of their testing partner and the stranger inside the scanner. Participants rated themselves as feeling significantly closer to the close other than to the stranger, indicating a good basis for exploring social distance effects on contagion, and a social distance effect was also found on learning and feedback processing (as discussed in section **7.3.2.**). Despite successfully increasing participants' belief in the experimental manipulation, whether or not participants believed that the other agents were real had no significant effect on the strength of contagion. This finding provides further support for the lack of a social basis for contagion in a TD task.

Whilst an effect of social distance has previously been found in an adolescent sample (Reiter et al., 2019), with significantly stronger contagion for the preferences of more risk-seeking peers, it is possible that this effect is specific to adolescent samples

and tasks assessing risk preferences. Previous research has also found that adolescents' risk perception was more influenced by the choices of own-age peers than non-peers (Knoll et al., 2015), and a later study by the same research group also found an own-age bias as well as a bias towards becoming more risk-seeking after viewing the choices of own-age peers in adolescents (Knoll et al., 2017). Future research could determine whether a social distance effect is also found in an adolescent sample in other discounting tasks such as TD or effort discounting.

7.3.5. Autism, contagion, and learning

In **chapter 3**, one sample of autistic participants, and a matched subset of NT participants were compared, and there were no significant group differences in D_{KL} or accuracy. There was also strong Bayesian support for equivalent contagion and accuracy between the two groups. This demonstrates that autistic participants were able to learn the preferences of others and were influenced by them in the same way as NT participants. Whilst the sample of autistic participants included for analysis in this chapter was small (N = 12), there was also no relationship between autistic traits and contagion, providing further support for this finding.

In **chapters 3** and **4**, the relationship between autistic traits and contagion in two NT samples was explored. In **chapter 3**, this data was collapsed across the two NT samples (i.e., the Study 1 and Study 2 samples), and separate regression analyses were run for impulsive (i.e., increasingly positive log *k*) and patient (i.e., increasingly negative log *k*) others, with subscale scores on the AQ (Baron-Cohen et al., 2001) entered as predictors of D_{KL}. The subscale scores of the AQ were not significantly predictive of contagion for more impulsive or more patient others. In **chapter 4**, separate analyses were run for the TD and PD tasks, with D_{KL} averaged across direction (i.e., more

positive/more negative). Once again, AQ subscale scores were not predictive of contagion for either task. These findings indicate that autistic traits in NT samples are not associated with contagion.

Whilst it is of interest that there is no association between autism and contagion, these findings cannot be attributed to a lack of differences in a social effect, as the findings throughout this thesis do not support contagion being a social effect. However, a social distance effect on learning and feedback processing was observed in **chapter 5**, indicating that there may be a social basis to learning about other agents. However, a group difference in accuracy was not observed between the NT and ASC samples, despite previous research finding differences in social PE signalling in the ACCg in autistic individuals (Balsters et al., 2017). Research has also provided support for differences in a social distance effect on discounting on behalf of others in a TD task in autistic participants, with steeper TD on behalf of close others in autistic participants versus NT participants (Warnell et al., 2019). Future research could explore whether a social distance effect is also found for learning the value preferences of others using real-world close and distant others, as in **chapter 5**.

Although the effect observed here does not appear to be a social effect, the finding of no association between autism, and autistic traits, and contagion and learning is of interest for decision-making and reward processing literature in autism. Previous research has found that autistic adults were more consistent and inflexible in their choices than NT controls in a PD task (Wu et al., 2018), yet participant's choices here changed to become more like that of the other. In a probability reversal learning task, autistic participants reverted back to previously preferred responses more quickly than NT controls (D'Cruz et al., 2013), indicating that autistic participants may change back to

their own preferences more quickly after adopting the preferences of another agent. Future research should explore whether the duration of the contagion effect differs across NT and ASC samples.

Predictive coding accounts of autism have found evidence for differences in forming precise internal prior beliefs and estimations of the likelihoods of new information, in both social and non-social contexts between groups of autistic and non-autistic participants (Lawson et al., 2017; Pellicano & Burr, 2012; Van De Cruys et al., 2014), and assessed on a continuum of autistic traits in NT populations (Karvelis et al., 2018). Despite these differences, the finding of no relationship between autistic traits and contagion here, and no group differences in contagion and learning of value preferences suggests a context in which Bayesian beliefs are accurately formed and updated.

7.4. Final conclusions

Taken together, the findings of this thesis show that contagion is a strong and reproducible effect, although the findings of this thesis were unable to index individual differences in this effect. Behavioural contagion effects were found in both NT and autistic adults, with no association between contagion and levels of autistic traits. No significant activation related to contagion of value preferences was found in the fMRI study, although future research could explore different methods of indexing contagion, such as neural oscillations, or patterns of connectivity.

Stronger contagion effects were found when using a Bayesian method of contagion which accounted for uncertainty, versus the shift measure employed in previous research (Garvert et al., 2015). However, these findings also indicate a bias towards becoming more patient in TD tasks, and more risk-seeking in PD tasks. Future

research could explore whether varying the extent to which the behaviour of the other agents differs from the participant impacts upon the contagion effect both for agents with increasingly negative discounting parameters (i.e., patient/risk-seeking) and increasingly positive discounting parameters (i.e., impulsive/risk-aversive).

Similarities and differences in contagion and discounting across TD and PD tasks were also identified, with correlations observed between contagion across tasks, and between discounting parameters for both tasks. However, participant's own choices were noisier in the PD task versus the TD task, and further research should explore the effects of participant uncertainty in their own choices on contagion. The findings throughout this thesis also indicate that contagion in discounting tasks may not be a social effect, as there were no differences in contagion dependent on the actual or believed identity of the other agent. Whilst there were no effects of social distance on contagion, participants were significantly more accurate at making choices on behalf of others that they were close to, and stronger feedback-related BOLD responses were observed in response to the feedback of close versus distant others, indicating that there was a social distance effect on learning. As there was no social basis found to the contagion effect here, the finding of no relationship between autism and contagion cannot tell us about social processes in autism, although future research should explore whether the same social distance effect is found on learning and feedback processing in discounting tasks in autistic participants.

References

Adolphs, R. (2008). The Social Brain: Neural basis of social knowledge. *Annual Review of Psychology, 60,* 693-716.

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*(4), 463-496.

Alexander, W. H., & Brown, J. W. (2010). Hyperbolically discounted temporal difference learning. *Neural Computation*, *22*(6), 1511–1527.

American Psychiatric Association. (2003) *Diagnostic and statistical manual of mental disorders* (5th ed.).

Amiez, C., Joseph, J. P., & Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *The European Journal of Neuroscience, 21*(12), 3447–52.

Amlung, M., Marsden, M., Holshausen, K., Morris, V., Patel, H., …, McCabe, R. E. (2019). Delay discounting as a transdiagnostic process in psychiatric disorders: A meta-analysis. *JAMA Psychiatry, 76*(11), 1176-1186.

Amlung, M., Vedelago, L., Acker, J., Balodis, I., & MacKillop, J. (2017). Steep delay discounting and addictive behavior: A meta-analysis of continuous associations. *Addiction,* 112(1), 51-61.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268–277.

Antrop, I., Stock, P., Verté, S., Wiersema, J. R., Baeyens D., & Roeyers, H. (2006). ADHD and delay aversion: The influence of non-temporal stimulation on choice for delayed rewards. *Journal of Child Psychiatry, 47*(11), 1152-1158.

Apps, M. A. J., Lockwood, P. L., & Balsters, J. H. (2013). The role of the midcingulate

cortex in monitoring others' decisions. *Frontiers in Neuroscience, 7*(7), 1–7.

Apps, M. A. J., & Ramnani, N. (2014). The anterior cingulate gyrus signals the net value of others' rewards. *Journal of Neuroscience, 34*(18), 6190-6200.

Apps, M. A. J., & Ramnani, N. (2017). Contributions of the medial prefrontal cortex to social influence in economic decision-making. *Cerebral Cortex, 29*(9), 4635-4648.

Apps, M. A. J., Rushworth, M. F. S., & Chang, S. W. C. (2016). The anterior cingulate gyrus and social cognition: Tracking the motivation of others. *Neuron, 90*, 692-707.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1–70.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences, 38*(2), 451-460.

Balsters, J. H., Apps, M. A. J., Bolis, D., Lehner, R., Gallagher, L., & Wenderoth, N. (2017). Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain, 140*(1), 235–246.

Balsters, J. H., & Ramnani, N. (2008). Symbolic representations of action in the human cerebellum. *NeuroImage, 43*(2), 388-398.

Balsters, J. H., & Ramnani, N. (2011). Cerebellar plasticity and the automation of first-order rules. *The Journal of Neuroscience, 31*(6), 2305–2312.

Balsters, J. H., Robertson, I. H., & Calhoun, V. D. (2013). BOLD frequency power indexes working memory performance. *Frontiers in Human Neuroscience, 7*, 207.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "Theory of Mind"? *Cognition, 21*, 37–46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M., & Behrens, T. E. J. (2021). Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron, 109*(4), 713-723.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427.

Basile, A. G., & Toplak, M. E. (2015). Four converging measures of temporal discounting and their relationships with intelligence, executive functions, thinking dispositions, and behavioral outcomes. *Frontiers in Psychology, 6*, 728.

Bathelt, J., Johnson, A., Zhang, M., & Astle, D. E. (2019). The cingulum as a marker of individual differences in neurocognitive development. *Science Reports, 9*, 2281.

Beer, J. S., & Ochsner, K. N. (2006). Social cognition: A multi-level analysis. *Brain Research, 1079*(1), 98-105.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008).

Associative learning of social value. *Nature, 456*, 245-249.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214–1221.

Bennett, D. (2015). The neural mechanisms of Bayesian belief updating. *The Journal of Neuroscience, 35*(50), 16300–16302.

Blackburn, M., & El-Deredy, W. (2013). The future is risky: Discounting of delayed and uncertain outcomes. *Behavioural Processes, 94*, 9–18.

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron, 62*, 733–743.

Botha, M., Hanlon, J., & Williams, G. L. (2021). Does language matter? Identity-first versus person-first language use in autism research: A response to Vivanti. *Journal of Autism and Developmental Disorders*, 1-9.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*(12), 539–546.

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective & Behavioral Neuroscience, 9*(1), 16–27.

Bowler, D. M., & Worley, K. (1994). Susceptibility to social influence in adults with Asperger's Syndrome: A research note. *Journal of Child Psychology and Psychiatry, 35*(4), 689–697.

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on

Pellicano and Burr. *Trends in Cognitive Sciences, 16,* 573–574.

Burke, C. J., Tobler, P. N., Schultz, W., & Baddeley, M. (2010). Striatal BOLD response reflects the impact of herd information on financial decisions. *Frontiers in Human Neuroscience, 4*, 48.

Bury, S. M., Jellett, R., Spoor, J. R., & Hedley, D. (2020). "It defines who I am" or "it's something I have": What language do [autistic] Australian adults [on the autism spectrum] prefer? *Journal of Autism and Developmental Disorders.* Advance online publication.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology, 20*(13), 1165–1170.

Carlisi, C. O., Norman, L., Murphy, C. M., Christakou, A., Chantiluke, K., Giampietro, V., … Rubia, K. (2017). Comparison of neural substrates of temporal discounting between youth with autism spectrum disorder and with obsessive-compulsive disorder. *Psychological Medicine, 47*(14), 2513-2527.

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2008). Perceived similarity and neural mirroring: Evidence from vicarious error processing. *Social Neuroscience, 4*(1), 85-96.

Chantiluke, K., Christakou, A., Murphy, C. M., Giampietro, V., Daly, E. M., Ecker, C., … Rubia, K. (2014). Disorder-specific functional abnormalities during temporal discounting in youth with attention deficit hyperactivity disorder (ADHD), autism and comorbid ADHD and autism. *Psychiatry Research, 223*(2), 113-120.

Chong, T. T. J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex.

Current Biology, 18, 1576-1580.

Chung, D., Christopoulos, G. I., King-Casas, B., Ball, S. B., & Chiu, P. H. (2015). Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. *Nature Neuroscience, 18*(6), 912–916.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*(1), 591–621.

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: A Bayesian framework for understanding models of psychosis. *Psychopharmacology*, *206*(4), 515–530.

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. (2009). Effort-based cost-benefit valuation and the human brain. *Journal of Neuroscience, 29*(14), 4531-4541.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgments according to its reliability. *The Journal of Neuroscience, 37*(25), 6066–6074.

Demetriou, L., Kowalczyk, O. S., Tyson, G., Bello, T., Newbould, R. D., & Wall, M. B. (2018). A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. *NeuroImage*, *176*, 404–416.

Delmonte, S., Balsters, J. H., McGrath, J., Fitzgerald, J., Brennan, S., Fagan, A. J., & Gallagher, L. (2012). Social and monetary reward processing in autism spectrum disorders. *Molecular Autism, 3*(1), 7.

Demurie, E., Roeyers, H., Baeyens, D., & Sonuga-Barke, E. (2012). Temporal discounting of monetary rewards in children and adolescents with ADHD and autism spectrum disorders. *Developmental Science*, *15*(6), 791-800.

Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., … Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Computational Biology*, *10*(9), e1003810.

Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., & Bodfish, J. W. (2011). Reward circuitry function in autism during face anticipation and outcomes. *Journal of Autism and Developmental Disorders, 42,* 147–160.

Dienes Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781.

Doya, K. (2007). Reinforcement learning: Computational theory and biological mechanisms. *HFSP Journal*, *1*(1), 30-40.

Eisenberg, D., Golberstein, E., Whitlock, J. L., & Downs, M. F. (2013). Social contagion of mental health: Evidence from college roommates. *Health Economics, 22*(8), 965–986.

Fan, Y., Duncan, N. W., Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 903-911.

Fareri, D. S., Niznikiewicz, M. A., Lee. V. K., & Delgado, M. R. (2012). Social network modulation of reward-related signals. *Journal of Neuroscience, 32*(26), 9045-9052.

Fletcher, P. C., & Frith, C. D. (2008) Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews*

Neuroscience 10, 48–58.

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience, 33*(1), 299–324.

Frank, L. R., Buxton, R. B., & Wong, E. C. (2001). Estimation of respiration-induced noise fluctuations from undersampled multislice fMRI data. *Magnetic Resonance in Medicine, 45*(4), 635–644.

Friston K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London: Biological sciences, 360*(1456), 815–836.

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind". *Cognition, 50,* 115-132.

Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J., & Dolan, R. J. (2015). Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron 85*(2), 418-428.

Glimcher, P. W., & Fehr, E. (2014). *Neuroeconomics: Decision making and the brain* (2nd Ed.), Elsevier Academic Press.

Göllner, L. M., Ballhausen, N., Kliegel, M., & Forstmeier, S. (2018). Delay of gratification, delay discounting and their associations with age, episodic future thinking, and future time perspective. *Frontiers in Psychology, 8,* 2304.

Gonzalez-Gadea, M. L., Sigman, M., Rattazzi, A., Lavin, C., Rivera-Rei, A., Marino, J., … Ibanez, A. (2016). Neural markers of social and monetary rewards in children with Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder. *Scientific Reports, 6*(1), 30588.

Goris, J., Brass, M. Cambier, C., Delplanque, J., Wiersema, J. R., & Braem, S. (2019). The relation between preference for predictability and autistic traits. *Autism Research, 13*(7), 1144-1154.

Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition, 25*(5), 715–723.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*(5), 769–792.

Grootoonk, S., Hutton, C., Ashburner, J., Howseman, A. M., Josephs, O., Rees, G., … Turner, R. (2000). Characterization and correction of interpolation effects in the realignment of fMRI time series. *NeuroImage, 11*(1), 49–57.

Hamilton, A. F. d. C. (2009). Research review: Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*(8), 881-892.

Harrigan, N., Achananuparp, P., & Lim, E. P. (2012). Influentials, novelty, and social contagion: The viral power of average friends, close communities, and old news. *Social Networks, 34*(4), 470–480.

Hayden, B. Y., & Platt, M. L. (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. *Current Biology, 17*(1), 49–53.

Heerey, E. A., Matveeva, T. M., & Gold, J. M. (2011). Imagining the future: Degraded representations of future rewards and events in schizophrenia. *Journal of Abnormal Psychology* 120, 483–489.

Henco, L., Brandi, M. L., Lahnakoski, J. M., Diaconescu, A. O., Mathys, C., & Schilbach, L. (2020). Bayesian modelling captures inter-individual differences in social belief computations in the putamen and insula. *Cortex, 131*, 221–236.

Henco, L., Diaconescu, A. O., Lahnakoski, J. M., Brandi, M. L., Hörmann, S., Hennings, J., … Mathys, C. (2020). Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder. *PLoS Computational Biology*, *16*(9), e1008162.

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences*, *20*(3), 204–213.

Hill, M. R., Boorman, E. D., & Fried, I. (2016). Observational learning computations in neurons of the human anterior cingulate cortex. *Nature Communications, 7*, 12722.

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science, 3*(302), 120-122.

Jackson, J. N. S., & MacKillop, J. (2016). Attention-deficit/hyperactivity disorder and monetary delay discounting: A meta-analysis of case-control studies. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1, 316–325.

Joiner, J., Piva, M., Turrin, C., & Chang, S. W. C. (2017). Social learning through prediction error in the brain. *NPJ Science of Learning, 2*(8), 1-8.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–33.

Kang, S. K., Hirsh, J. B., & Chasteen, A. L. (2010). Your mistakes are mine: Self-other overlap predicts neural response to observed errors. *Journal of Experimental Social Psychology*, *46*(1), 229-232.

Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriè, O. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife, 7,* 1-29.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kekic, M., McClelland, J., Bartholdy, S., Chamali, R., Campbell, I. C., & Schmidt, U. (2020). Bad things come to those who do not wait: Temporal discounting is associated with compulsive overeating, eating disorder psychopathology and food addiction. *Frontiers in Psychiatry, 10,* 978.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism: The International Journal of Research and Practice*, *20*(4), 442–462.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25–41.

Kirby, K. N. (2009) One-year temporal stability of delay-discount rates. *Psychonomic Bulletin & Review* 16, 457–462.

Kirby, K. N., Winston, G. C., & Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences, 15,* 213–222.

Klein-Flügge, M. C., Hunt, L.T., Bach, D. R., Dolan, R. J., & Behrens, T. E. J. (2011). Dissociable reward and timing signals in human midbrain and ventral striatum. *Neuron, 72*(4), 654-664.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009).

Reinforcement learning signal predicts social conformity. *Neuron, 61*(1), 140–151.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719.

Knolle-Veentjer, S., Huth, V., Ferstl, R., Aldenhoff, J. B., & Hinze-Selch, D. (2008). Delay of gratification and executive performance in individuals with schizophrenia: Putative role for eating behavior and body weight regulation. *Journal of Psychiatric Research, 42*(2), 98–105.

Knoll, L. J., Leung, J. T., Foulkes, L., & Blakemore, S. J. (2017). Age-related differences in social influence on risk perception depend on the direction of influence. *Journal of Adolescence, 60*, 53–63.

Knoll, L. J., Magis-Weinberg, L., Speekenbrink, M., & Blakemore, S. J. (2015). Social influence on risk perception during adolescence. *Psychological Science, 26*(5), 583–592.

Kohls, G., Schulte-Rüther, M., Nehrkorn, B., Müller, K., Fink, G. R., Kamp-Becker, I., … Konrad, K. (2013). Reward system dysfunction in autism spectrum disorders. *Social Cognitive and Affective Neuroscience, 8*, 565–572.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence or common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502.

Lamm, C., Fischer, M. H., & Decety, J. (2007). Predicting the actions of others taps into one's own somatosensory representations - A functional MRI study. *Neuropsychologia*, *45*(11), 2480–2491.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*(9), 1293–1299.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience, 8*(302), 1-10.

Lazzaro, S. C., Weidinger, L., Cooper, R. A., Baron-Cohen, S., Moutsiana, C., & Sharot, T. (2018). Social conformity in autism. *Journal of Autism and Developmental Disorders.*

Lee, V. K., & Harris, L. T. (2013). How social cognition can inform social decision making. *Frontiers in Neuroscience, 7*, 259.

Lempert, K., Steinglass, J., Pinto, A., Kable, J., & Simpson, H. (2019). Can delay discounting deliver on the promise of RDoC? *Psychological Medicine, 49*(2), 190-199.

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.

Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience, 31*(14), 5504-5511.

Lin, A., Rangel, A., & Adolphs, R. (2012). Impaired learning of social compared to monetary rewards in autism. *Frontiers in Neuroscience, 6*(143), 1–7.

Lockwood, P. L. (2016). The anatomy of empathy: Vicarious experience and disorders of social cognition. *Behavioural Brain Research, 311*, 255-266.

Lockwood, P. L., Apps, M. A. J., & Chang, S. W. C. (2020). Is there a 'social' brain?

Implementations and algorithms. *Trends in Cognitive Sciences, 24*(10), 802-813.

Lockwood, P. L., Apps, M. A. J., Roiser, J. P., & Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *Journal of Neuroscience, 35*(40) 13720-13727.

Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E., & Roiser, J. P. (2016). Neurocomputational mechanisms of prosocial learning and links to empathy. *PNAS, 113*(35), 9763-9768.

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications, 9*(1), 4747.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*(1), 93–120.

McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J. C., & Stout, S. C. (2009). A comparison of four models of delay discounting in humans. *Behavioural Processes, 81*(2), 256–9.

Meyer, M. L., Zhao, Z., & Tamir, D. I. (2019). Simulating other people changes the self. *Journal of Experimental Psychology: General, 148*(11), 1898-1913.

Mischel, W., Ebbesen, E. B., & Zeiss, A. R. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, *21*(2), 204–218.

Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of inter-temporal preferences. *PLoS Computational Biology, 12*(7), 1–17.

Myerson, J., Green, L., Scott Hanson, J., Holt, D. D., & Estle, S. J. (2003). Discounting

delayed and probabilistic rewards: Processes and traits. *Journal of Economic Psychology, 24*(5), 619–635.

Myerson, J., Green, L., & Morris, J. (2011). Modeling the effect of reward amount on probability discounting. *Journal of the Experimental Analysis of Behavior*, *95*(2), 175–187.

Nagel, I. E., Preuschhof, C., Li, S. C., Nyberg, L., Backman, L., Lindenberger, U., & Heekeren, H. R. (2011). Load modulation of BOLD response and connectivity predicts working memory performance in younger and older adults. *Journal of Cognitive Neuroscience, 23,* 2030–2045.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114–21.

Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of Mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology, 125*(6), 818–823.

O'Connell, G., Christakou, A., Haffey, A. T., & Chakrabarti, B. (2013). The role of empathy in choosing rewards from another's perspective. *Frontiers in Human Neuroscience, 7,* 174.

O'Connell, G., Hsu, C. T., Christakou, A., & Chakrabarti, B. (2017). Thinking about others and the future: Neural correlates of perspective taking relate to preferences for delayed rewards. *Cognitive, Affective, & Behavioral Neuroscience,* 1–8.

Odum, A. L., Becker, R. J., Haynes, J. M., Galizio, A., Frye, C., Downey, H., … Perez, D. M. (2020). Delay discounting of different outcomes: Review and

theory. *Journal of the Experimental Analysis of Behavior, 113*(3), 657–679.

Ogunlade, J. O. (1979). Personality characteristics related to susceptibility to behavioral contagion. *Social Behavior and Personality: An International Journal, 7*(2), 205–208.

Ohmura, Y., Takahasi, T., Kitamura, N., & Wehr, P. (2006). Three-month stability of delay and probability discounting measures. *Experimental and Clinical Psychopharmacology, 14*(3), 318-328.

O'Reilly, J. X., Jbabdi, S., & Behrens, T. E. J. (2012). How can a Bayesian approach inform neuroscience? *European Journal of Neuroscience, 35*(7), 1169–1179.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin, 143,* 521–542.

Palmer, C. J., Seth, A. K., & Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. *Conscious Cognition, 36,*376-389.

Paulsen, D. J., Platt, M. L., Huettel, S. A., & Brannon, E. M. (2012). From risk-seeking to risk-averse: The development of economic risk preference from childhood to adulthood. *Frontiers in Psychology, 3,* 313.

Park, S. A., Goïame, S., O'Connor, D. A., & Dreher, J. C. (2017). Integration of individual and social information for decision-making in groups of different sizes. *PLoS Biology, 15*(6), 3–28.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16,* 504–510.

Perreault, C., Moya, C., & Boyd, R. (2012). A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior, 33,* 449–459.

Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron*, 35, 975–987.

Peters, J., & Büchel, C. (2009). Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. *The Journal of Neuroscience, 29*(50), 15727–34.

Peterson, C. C., Slaughter, V., Peterson, J., & Premack, D. (2013). Children with autism can track others' beliefs in a competitive game. *Developmental Science, 16*(3), 443-450.

Powers, K. E., Somerville, L. H., Kelley, W. M., & Heatherton, T. F. (2016). Striatal associative learning signals are tuned to in-groups. *Journal of Cognitive Neuroscience, 28*(9), 1243-1254.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515.

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry, 18*(1), 178.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, *55*(2), 233–244.

Reiter, A. M. F., Suzuki, S., O'Doherty, J. P., Li, S-C., & Eppinger, B. (2019). Risk contagion by peers affects learning and decision-making in adolescents. J*ournal of Experimental Psychology: General, 148*(9), 1494–1504.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In

Black, A. H., & Prokosy W. F. *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks, 15*(4-6), 507–521.

Ruff, C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience, 15,* 549–562.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342(6157), 482-484.*

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in cognitive sciences, 19*(2), 65–72.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology, 57*(1), 87–115.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–9.

Schultz, W., & Dickinson, A. (2000). Neural coding of prediction errors. *Annual Review of Neuroscience, 23*, 473-500.

Schurz, M., & Perner, J. (2015). An evaluation of neurocognitive models of theory of mind. *Frontiers in Psychology, 6*, 1610.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9–34.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883–5.

Sevgi, M., Diaconescu, A. O., Henco, L., Tittgemeyer, M., & Schilbach, L. (2020). Social Bayes: Using Bayesian modeling to study autistic trait–related differences in social cognition. *Biological Psychiatry, 87*(2), 185-193.

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., & Gvirts, H. Z. (2019). Herding brains: A core neural mechanism for social alignment. *Trends in Cognitive Sciences, 23*(3), 174-186.

Stephan, K. E., Baldeweg, T., & Friston, K. J. (2006). Synaptic plasticity and dysconnection in schizophrenia. *Biological Psychiatry, 59*(10), 929–939.

Stevenson, J. L., & Hart, K. R. (2017). Psychometric properties of the autism spectrum quotient for assessing low and high levels of autistic traits in college students. *Journal of Autism and Developmental Disorders, 47*(6), 1838-1853.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, Massachusetts: MIT press.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., … Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron, 74*, 1125-1137.

Suzuki, S., Jensen, E. L. S., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioral contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *PNAS, 113*(14), 3755–3760.

Todd, N., Moeller, S., Auerbach, E. J., Yacoub, E., Flandin, G., & Weiskopf, N. (2016). Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts. *NeuroImage, 124*(Pt A), 32–42.

Tomlin, D., Nedic, A., Prentice, D., Holmes, P., & Cohen, J. D. (2013). The neural

substrates of social influence on decision making. *PLoS ONE, 8*(1), 1-11.

Thomas, L., Lockwood, P. L., Garvert, M. M., & Balsters, J. H. (2021). Contagion of temporal discounting value preferences in neurotypical and autistic adults. *Journal of Autism and Developmental Disorders.* Advance online publication.

Tunçgenç, B., El Zein, M., Sulik, J., Newson, M., Zhao, Y., Dezecache, G., & Deroy, O. (2021). Social influence matters: We follow pandemic guidelines most when our close circle does. *British Journal of Psychology.* Advance online publication.

Van De Cruys, S., Evers, K., Van Der Hallen, R., Van Eylen, L., Boets, B., De-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121,* 649-675.

Van Hoorn, J., Van Dijk, E., Crone, E. A., Stockmann, L., & Rieffe, C. (2017). Peers influence prosocial behavior in adolescent males with autism spectrum disorders. *Journal of Autism and Developmental Disorder, 47*(7), 2225-2237.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30*(3), 829–858.

Vilares, I., & Kording, K. (2011) Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39.

Warnell, K. R., Maniscalco, S., Baker, S., Yi, R., & Redcay, E. (2019). Social and delay discounting in autism spectrum disorder. *Autism Research*, *12*(6), 870–877.

Wheeler, L. (1966). Toward a theory of behavioral contagion. *Psychological Review*, *73*(2), 179–192.

Wulfert, E., Block, J. A., Santa Ana, E., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality, 70*(4), 533–552.

Yafai, A. F., Verrier, D., & Reidy, L. (2014). Social conformity and autism spectrum disorder: A child-friendly take on a classic study. *Autism, 18*(8), 1007–1013.

Yu, L. Q., Lee, S., Katchmar, N., Satterthwaite, T. D., Kable, J. W., & Wolf, D. H. (2017). Steeper discounting of delayed rewards in schizophrenia but not first-degree relatives. *Psychiatry Research* 252, 303–309.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science, 22*(7), 894–900.

Zheng, H., & Zhu, L. (2013). Neural mechanism of proposer's decision-making in the ultimatum and dictator games. *Neural Regeneration Research, 8*(4), 357–62.

Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems, 93*, 77-87.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavia, 67*(6), 361-370.