UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Cognitive & Affective Modulators of Prosocial Behavior

Permalink

https://escholarship.org/uc/item/06r9r3sv

Author

Arroyo, Lucila

Publication Date

2023

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Cognitive & Affective Modulators of Prosocial Behavior

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Social Sciences

by

Lucila Arroyo

Dissertation Committee:
Associate Professor Mimi Liljeholm, Chair
Professor emerita Linda J. Levine
Professor Michael McBride

2023

© 2023 Lucila Arroyo

DEDICATION

To René and Kimi, for being my light

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Professor Mimi Liljeholm, for the thorough guidance in the making of this dissertation and for helping my career move forward. Thank you for believing in me.

To my partner, René Zamarripa, thank you for the help with this work, but, above this, to you and Kimi, thank you for helping me in our personal lives and for giving me life. To Zuri, thank you for the smiles.

I want to thank Professor Louis Narens for accepting me into the Ph.D. program.

I want to thank my best friend, Delfina, for helping me when I got stuck in my own mind.

I want to thank my parents for supporting me in this journey. And I want to thank them and my sister for being there when I need them.

I want to thank Enriqueta for the help at the beginning of this year.

I would like to thank committee members, Professor Linda J. Levine and Professor Michael McBride, for their feedback on the dissertation.

I want to thank Dr. Kimberly Jameson for checking in on me.

Portions of Chapter 1 of this dissertation are a reprint of the material as it appears in Socio-Affective Traits Mediating Charitable Giving, published in Proceedings of the Annual Meeting of the Cognitive Science Society, used with permission from the Cognitive Science Society. The co-author listed in this publication, Mimi Liljeholm, directed and supervised research which forms the basis for the dissertation.

This work was supported by CAREER grant 1654187 and grant 1844632 from the National Science Foundation awarded to Mimi Liljeholm.

VITA

Lucila Arroyo

2015 B.A. in Economics, University of CEMA

2015 Undergraduate Student Exchange Program, San Diego State University

2023 Ph.D. in Mathematical Behavioral Sciences, University of California, Irvine

FIELD OF STUDY

Decision making; emotion; psychology and behavioral economics.

PUBLICATIONS

Arroyo, L., & Liljeholm, M. (2023). Socio-Affective Traits Mediating Charitable Giving. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

ABSTRACT OF THE DISSERTATION

Cognitive & Affective Modulators of Prosocial Behavior

by

Lucila Arroyo

Doctor of Philosophy in Social Sciences

University of California, Irvine, 2023

Professor Mimi Liljeholm, Chair

In 2022, charitable giving accounted for roughly 2% of the gross domestic product in the U.S., with 64% of donations being made by individuals. Despite the significant impact of this industry on the economy, little is known about the socio-affective constructs that mediate charitable giving. In this dissertation, I assess the influences of a range of motivational and cognitive states and traits on real-world charity donations. In Chapter 1, I demonstrate that individual differences in narcissism and approval-seeking predict decisions to make public vs. anonymous donations, and that these relationships are modulated by social information about the decisions made by other donors. In Chapter 2, I use a real-time interactive "advisor-decider" task, in which advice given by one participant results in an onerous workload for another participant, to show that self-conscious affect based on performance in one domain shapes decisions to engage in charitable giving in an unrelated domain. Advisors that performed at or worse than the norm, in terms of giving incorrect advice, made more frequent subsequent charity donations: Intriguingly, when advisors were given social information about their performance relative to the norm, this pattern was reversed, such that advisors that performed worse than the norm made less frequent donations. Finally, in

Chapter 3, I explored whether a recently demonstrated preference for instrumental divergence – the degree to which voluntary actions differ with respect to their outcome distributions – depends on whether monetary decision outcomes are kept for oneself or donated to a charity of one's choice. I found that the intrinsic motivation associated with greater instrumental control applies selectively to decisions that earn money for oneself – no such preference was observed for decisions that earned money for charities. Collectively, these studies serve to better characterize the social, cognitive, and motivational processes that mediate prosocial behavior.

INTRODUCTION

Helping involves any action that has as consequence the provision of a benefit to another person (Dovidio et al., 2017). Such a broad definition implies that helping involves a wide range of diverse behaviors, from holding the door open for someone carrying heavy bags, to spending every summer providing free tutoring services for children without access to education. Due to how qualitatively different types of helping can be, Pearce & Amato (1980) proposed a taxonomy of helping that categorizes helping forms into three dimensions: planned, formal vs. spontaneous, informal help; serious vs. non-serious help; and giving, indirect vs. doing, direct help. The current work is interested in cognitive and affective aspects behind a specific type of helping, in this specific case planned, serious, and involving a combination of giving and doing: donations to charities.

Chapman et al. (2022) proposed a theoretical framework they termed "Charitable Triad". They argued that the characteristics of three actors – donors, beneficiaries, and fundraisers –, as well as the interactions between their characteristics, influence the donations to charities. For example, people who are higher in empathy are in general more likely to be donors, beneficiaries who are in worse conditions tend to receive more support, and fundraising organizations with greater impact on the cause are seen as more legitimate (Chapman et al., 2022). The current work focused on the donors, whose decisions are shaped by both emotion and cognition (Zagefka & James, 2015). The present work assessed the influences of a range of motivational and cognitive states and traits on charity donations.

In Chapter 1 I focused on motivations that are self-serving. I assessed how social approval-seeking and narcissism shape decisions to donate, and how information about others'

donation decisions modulates those relationships. In Chapter 2, I focused on self-conscious emotions. I used an interpersonal guilt induction task to assess how self-conscious emotions in one domain influence charitable giving in another domain, and I again looked at the role of social information. Finally, in Chapter 3, I explored whether a recently demonstrated preference for instrumental divergence – the degree to which voluntary actions differ with respect to their outcome distributions – depends on whether monetary decision outcomes are kept for oneself or donated to a charity of one's choice.

Chapter 1: Social Approval-Seeking & Narcissism

1.1 Background & Significance

Charitability is often assumed to be motivated by compassion and empathy (e.g., Lim & DeSteno, 2016); however, egocentric traits, like approval-seeking and narcissism may provide self-serving incentives for prosocial decisions. In this chapter, I present a study aimed at exploring prosocial behavior that is mainly driven by an individual's own wants and needs rather than the wants and needs of others. To that end, three well-validated dimensions of self-serving behavior are assessed in participants making decisions about making public or anonymous charitable donations.

1.1.1 Simulated Compassion

Genuine compassion has been defined as "the feeling that arises in witnessing another's suffering and that motivates a subsequent desire to help" (Goetz, Keltner, & Simon-Thomas, 2010). It involves care, approach, and prosocial motivation (Singer & Klimecki, 2014). In contrast, Catarino et al. (2014) define submissive (henceforth *simulated*) compassion as caring that serves "self-advancing or protective needs, such as wanting to please others, to be liked or thought well of, and to avoid rejection". They found that caring shame – the fear of being criticized for not being caring enough – and self-image goals predicted simulated compassion. Moreover, simulated compassion was highly correlated with caring guilt – focused on regret and a sense of responsibility – as well as submissive behavior, anxiety, and

stress. Here, I assess whether simulated compassion can predict prosocial behavior, in terms of the probability of engaging in a prosocial act and a quantitative measure of the level of help.

1.1.2 Social Anxiety

Social anxiety involves a fear of being evaluated by others, including a fear of rejection, and it can occur either when someone is currently being evaluated or when there is a possibility of being evaluated (Leary & Kowalski, 1997; Teachman & Allen, 2007). Łakuta (2018) identifies five dimensions of social anxiety, namely: negative view of the self as a social object, self-focused attention (e.g., constant thinking about how one looks or sounds to others), safety behaviors, somatic and cognitive symptoms (e.g., sweating and mental blanks), and anticipatory and post-event rumination. Weisman et al. (2011) found social anxiety to be related to behaving submissively as well as perceiving oneself as having a low social rank. Here, I investigate how the fear of social rejection – seen in both simulated compassion and social anxiety – influences the decision to donate publicly vs. anonymously.

1.1.3 Narcissism

Narcissism can be divided into two broad categories: grandiose and vulnerable. Grandiose narcissism includes traits such as self-enhancement, entitlement, and dominance; while vulnerable narcissism includes entitlement, distrust of others, and defensive grandiosity to obscure feelings of inadequacy (Miller et al., 2013; Miller et al., 2017). While the first is associated with inflated self-esteem, the latter is associated with fragile self-esteem (Ackerman et al., 2011).

In the present work, particular attention is paid to aspects of narcissism measured by the Entitlement/Exploitativeness (EE) assay, which measures entitled beliefs and behaviors in interpersonal contexts, including a willingness to manipulate others (Ackerman et al., 2011; Gentile et al., 2013). EE narcissism has been argued to relate to both grandiose and vulnerable narcissism (Ackerman et al., 2011; Gentile et al., 2013).

Narcissism is usually related to negative psychosocial consequences, such as interpersonal difficulties and antisocial tactics – e.g., lying – (Muris et al., 2017). Here, I investigate whether EE Narcissism can promote prosocial behavior.

1.1.4 Social Conformity

In a field experiment on charity donations, Alpizar, Carlsson, & Johansson-Stenman (2008) investigated the effect of providing information about the typical dollar amount ($2, $5, or $10) contribution made by others, and found that in all cases the most common contribution aligned with the one provided as reference. Similarly, assessing the impact of multiple earlier donations on the donation of a subsequent donor, Sasaki (2019) found that the greater the number of similar donations among earlier donations, the greater the likelihood that a donor would match that modal donation. Here, it was assessed whether social information might interact with the three aforementioned measures in the decisions to donate to charity.

1.1.5 A Study on Self-Serving Incentives of Charitability

The present study investigated how self-serving motivations might promote donations of time and effort to gain money for charity when donations could be either Public or Anonymous and when participants received vs. did not receive information regarding other

people's decisions. I predicted that Simulated Compassion and Social Anxiety would increase both the probability of donating and the probability of donating publicly vs. anonymously when there was no information about other people's decisions. However, I expected Social Anxiety to interact with Social Information and influence participants' decisions to make their donations anonymous vs. public. Moreover, I expected EE Narcissism to reduce the probability of making donations anonymously, but to increase the probability of making a donation. Finally, I expected all three motivations to have a negative effect on the donation amount.

1.2 Methods

1.2.1 Participants

Two hundred and ten undergraduates (182 female, mean age 20.9 ± 3.22) at the University of California, Irvine (UCI) participated in the study for extra course credit. The study was posted to a cloud-based participant pool management system where any UCI student enrolled in a course that allows extra credit to be earned via research participation could sign up. The sample size was based on Gilbert et al. (2017b). Participants were compensated with course credit for the thirty minutes it took to complete the main tasks, but were not compensated for the time they decided to donate to gain money for a charity. All participants gave informed consent, and the Institutional Review Board of the institution approved the study. Participants that did not follow instructions in at least eighty percent of the rounds of

the main tasks were removed from analysis. Data analysis was performed on the remaining one hundred and eighty participants (157 female, mean age 20.8 ± 2.94).

1.2.2 Tasks

Charity Ratings Task Participants were presented with sixty real charities, one at a time. For each charity, they were shown its name, a picture —from the charity's website —, and its mission. They had to rate each charity by how deserving of assistance they believed it to be (*deservingness*) and how likely it was that they or someone they knew would directly benefit from its mission (*closeness*). Both ratings were given on a 7-point Likert scale ranging from "Not at all deserving" to "Extremely deserving" and "Not at all likely" to "Extremely likely", respectively. This task was based on Hare et al. (2010). A screen from the task is illustrated in Figure 1.1.

Donations Decisions Task Participants were presented with the same sixty charities and asked to submit a donation decision for each of them. They knew that one of those sixty donation decisions would be randomly selected to be performed. The donation decision consisted of (1) the number of Slider Tasks they committed to perform for the charity (see next subheading) and (2) whether they wanted the donation to be Anonymous or Public. Participants could commit to performing any number of Slider Tasks between 0 (No donation) and 20. For each Slider Task they committed to performing they could earn $1 for the respective charity. Each potential (positive) donation could be either Anonymous or Public. Anonymous donations would not be associated with the participant's name but had a 50% chance of being doubled by the researchers. Public donations had no possibility of being doubled but, if large enough,[1] the participant's name would be included on the "Donors

of the Week" list on a public website created for the study.[2] The link to the website was included in the recruitment message in the participant pool management system and was also emailed to all participants that completed this study. For each positive donation decision (1 to 20 Slider Tasks) participants had to select Anonymous or Public; when the donation decision was 0 Slider Tasks, they were instructed to select "N/A" (Not Applicable).

To assess susceptibility to social norms, participants were randomly assigned to a "Social Information" group (n=87), in which, for each donation decision, information was provided regarding what percentage of donations to that charity were Public vs. Anonymous in a pilot study (n=61). Critically, Public donations were much less frequent than Anonymous donations, so that, if susceptible to the donation decisions of their peers, participants should reduce their Public donations. A screen from the task is illustrated in Figure 1.2.

[1] A 'large enough' donation was $10 or larger, but participants were not specified this threshold.

[2] https://sites.google.com/view/thecharityproject/home

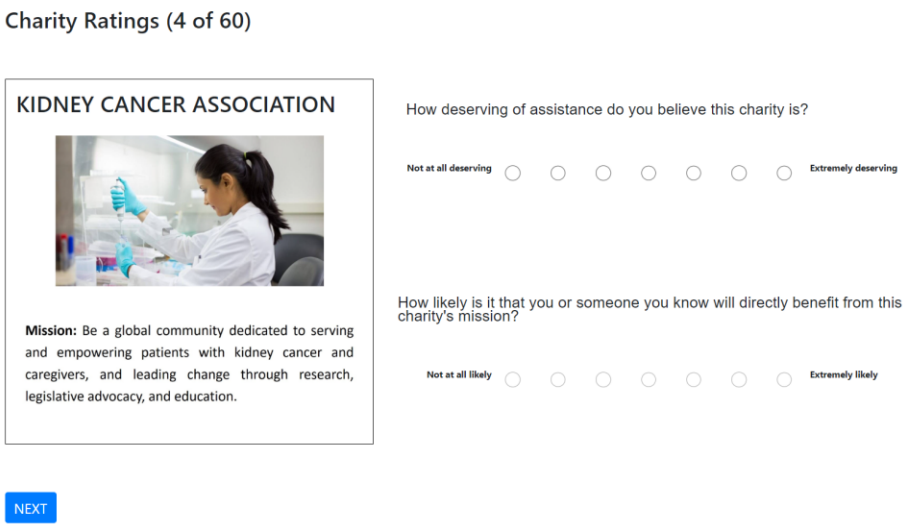

Figure 1.1 A Round of the Charity Ratings Task. On the left side of the screen participants saw the charity's name, picture, and mission. On the right side of the screen, participants had to respond: (i) "How deserving of assistance do you believe this charity is?" (ii) "How likely is it that you or someone you know will directly benefit from this charity's mission?".

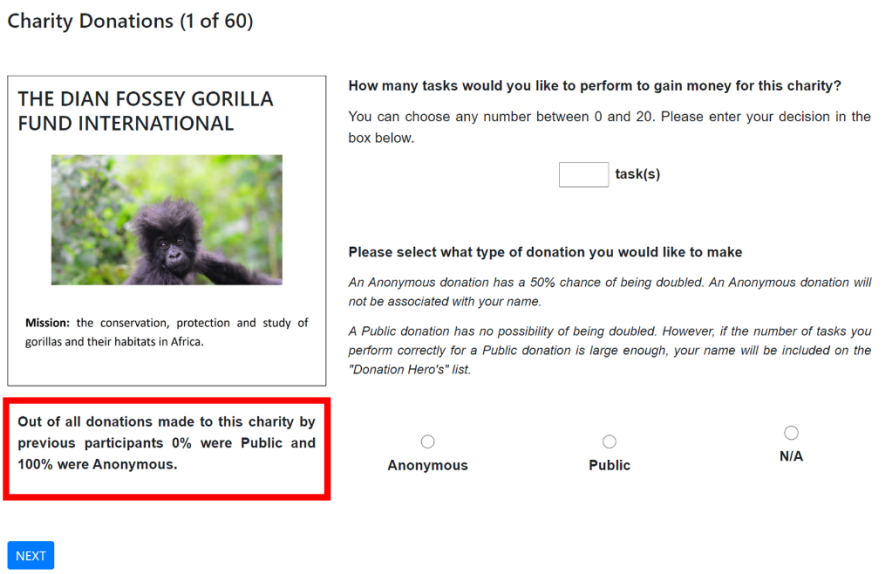

Figure 1.2 A Round of the Donation Decisions Task. On the left side of the screen participants saw the charity's name, picture, and mission. Participants in the Social Information Group were also shown a box (outlined in red in the figure) stating what percentage, out of all donations made to the charity by previous participants, was Public and what (complementary) percentage was Anonymous. On the right side of the screen, participants had to respond: (i) How many Slider Tasks they would like to perform to gain money for the charity and (ii) What type of donation they would like to make by selecting Public, Anonymous, or N/A (if choosing No donation). The order of the radio buttons was random on each round.

Slider Tasks Participants were presented with five sliders positioned randomly across the screen. All sliders began with their value set at zero (left of the slider), and participants had 30 seconds to move all values to fifty (middle of the slider). If, when the timer ran out, all sliders had a value of fifty, the participant would earn $1 for the selected charity. Participants could complete up to 20 Slider Tasks, with the position of the sliders on the screen being random at the beginning of each task. A Slider Task is illustrated in Figure 1.3.

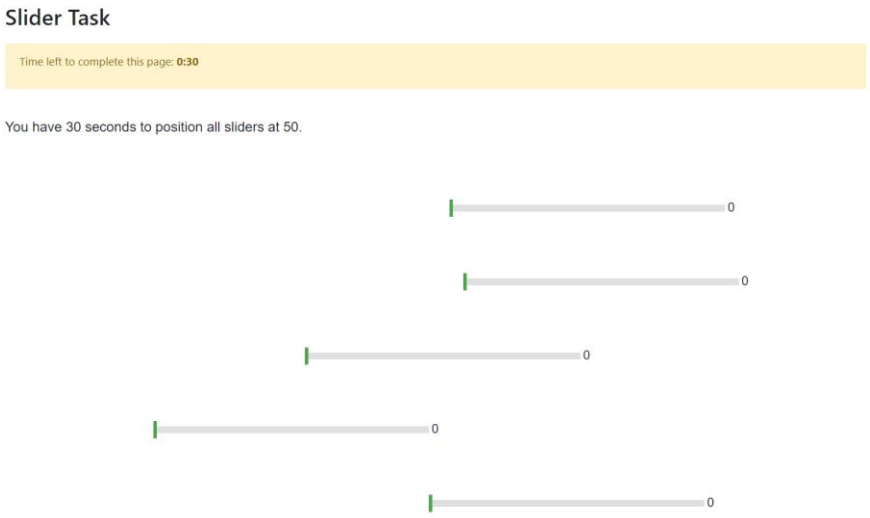

Figure 1.3 A Slider Task. In a Slider Task, participants were presented with five sliders randomly positioned across the screen and had 30 seconds to change all slider values from 0 to 50. The position in which the sliders appeared on the screen changed in every round.

1.2.3 Procedure

At the start of the experiment participants were instructed that the study had two phases. They needed to complete phase one to be compensated with extra course credit. Phase two

was optional and they would not receive compensation for completing it, instead, they would be donating their time and effort to gain money for a charity. During phase one participants filled out demographic information, completed the Warwick-Edinburgh Mental Well-being Scale, the Charity Rating Task, the Donation Decisions Task, the Warwick-Edinburgh Mental Well-being Scale for a second time, the Compassionate Engagement and Action Scale – Compassion to Others, and, lastly, the three self-report measures of interest (see below). Before making their donations decisions participants did a trial round of a Slider Task to learn about the task before making their commitments. At the end of phase one participants were told which donation decision was chosen to be performed. During phase two, participants were asked to complete their donation decision – that is, the Slider Tasks they had committed to. The total amount of money gained for the charity was determined by the number of Slider Tasks they completed correctly in phase two, with a fifty percent chance of doubling the total if the donation decision was Anonymous. Donations were sent to the corresponding charities.

1.2.4 Self-Report Measures

Submissive (Simulated) Compassion Scale (Catarino et al., 2014) (SCS) This scale assesses the extent to which an individual's compassionate acts are guided by simulated compassion. The scale consists of 10 statements regarding reasons for being caring, and participants responded on a five-point scale ranging from "Not at all like me" to "Extremely like me". Scores can range from 0 to 40.

Social Anxiety Questionnaire (Łakuta, 2018) (SAQ) This scale measures social anxiety defined as "a marked and persistent fear of negative evaluation in social situations". The

scale consists of 10 statements and participants responded on a five-point scale ranging from "Strongly Disagree" to "Strongly Agree". Scores can range from 10 to 50.

Narcissistic Personality Inventory – 13 (Gentile et al., 2013) (NPI) This scale is a brief measure of narcissism that provided a total score and three subscale scores. The scale consists of 13 pairs of attributes, for each pair participants had to choose the one that they most agreed with. Of interest was the subscale of Entitlement/Exploitativeness (NPI EE). The NPI EE scores can range from 0 to 4.

The Compassionate Engagement and Action Scale – Compassion to Others (Gilbert et al. (2017a) This scale measures the ability to be compassionate to others, it includes compassion attributes and specific compassionate actions. The scale consists of 13 items, including 3 filler items. Participants responded on a scale ranging from "Never" (1) to "Always" (10). Scores can range from 10 to 100.

Warwick-Edinburgh Mental Well-being Scale (Tennant et al., 2007) This scale measures mental wellbeing. The scale has 14 items and participants responded on a scale ranging from "None of the time" (1) to "All of the time" (5). Scores can range from 14 to 70.

1.2.5 Statistical Analyses

The influence of each construct of interest (i.e., SCS, SAQ, & NPI EE – including their interaction with social information), on (i) the proportion of public, (ii) anonymous, (iii) no donations, and (iv) average magnitude of donation was assessed using linear regressions, adding average deservingness and closeness ratings as covariates due to their expected influence on donation decisions (Hare et al., 2010). Unstandardized coefficients, t-statistics

and p-values are reported for significant results. Statistical analyses were implemented in JASP.

1.3 Results

The percentage of Slider Tasks completed correctly out of the ones committed to perform by participants was 88.73.

Public Donations Of the regressors of interest the only one that approached, but did not reach, significance was Simulated Compassion (SCS: $\beta=0.005$, $t(171)=1.703$, $p=0.09$). Apparently higher Public donations were observed in participants with higher SCS. Deservingness and Closeness were significant ($\beta=-0.041$, $t(171)=-2.466$, $p<0.02$; $\beta=0.025$, $t(171)=2.348$, $p=0.02$; respectively), such that Deservingness decreased and Closeness increased Public donations.

Anonymous Donations A higher score on the Narcissism – Entitlement/Explotativeness (NPI-EE) scale decreased the proportion of Anonymous donations ($\beta=-0.054$, $t(171)=-2.091$ $p<0.04$), as did Closeness ratings ($\beta=-0.059$, $t(171)=-3.590$, $p<0.001$).

No Donations SCS predicted the proportion of no donations ($\beta=-0.010$, $t(171)=-2.488$, $p<0.02$), that is, a higher SCS score increased the probability of making a donation. However, SCS interacted with Social Information, such that, for participants who received social information, the effect was counteracted ($\beta=0.012$, $t(171)=2.066$, $p=0.04$). Social Anxiety (SAQ) had a marginally significant interaction with Social Information; for participants in the Social Information group, a higher SAQ score increased the probability of donation ($\beta=-$

0.011, t(171)=-1.890, p=0.061). To continue, a higher NPI EE score decreased the probability of making a donation, but the effect did not reach significance (β=0.042, t(171)=1.673, p=0.096). Surprisingly, closeness also predicted a higher probability of not donating (β=0.035, t(171)=2.176, p=0.031).

Average Donation Amount SCS predicted a decrease in the average donation amount (β=-0.178, t(171)=-2.153, p<0.04). Moreover, participants in the Social Information group tended to donate a lower amount (β=-6.396, t(171)=-1.966, p=0.051). Finally, Deservingness predicted an increase in donation amount (β=1.159, t(171)=2.293, p<0.03).

1.4 Discussion

This study explored the influence of self-serving traits on charitable giving, using an online task in which participants could choose to exert time and effort that was subsequently translated into monetary donations. Participants had the option of making *Public Donations*, associated with the possibility of having one's name displayed on a "Donors of the Week" webpage; *Anonymous Donations*, not associated with any publicity, but with the possibility of the experimenters doubling the donated amount; or *No Donations*.

Simulated Compassion showed a tendency to increase the proportion of Public donations, but the effect was not significant. The focus of simulated compassion is wanting to be liked or thought well of (Catarino et al., 2014), which cannot be accomplished if other people do not know about one's actions, so this effect was expected. The lack of a significant effect

might be due to Public donations not having great dissemination, a limitation of the present study. However, the result provides preliminary evidence of the influence of this type of compassion on anonymous vs. public donations. Another, in this case significant, influence of Simulated Compassion was on the probability of donating. Higher Simulated Compassion predicted a higher probability of donating; however, this effect was different for participants who received Social Information from those who did not. Once participants learned that most people chose to make donations anonymously, the incentive to donate was weakened. If most people choose to make donations anonymously, then making a public donation might not provide the social acceptance that is sought; what is more, donations that are anonymous defeat the purpose of simulated compassion if no one finds out, so not donating might become a better option. Interestingly, Simulated Compassion also predicted a decrease in donation amount, indicating a lower limit in willingness to help.

Social Anxiety was found to have an influence on the probability of donating, differentially for the group of participants receiving vs. not receiving Social Information. The group that received Social Information had a higher probability of donating. Social anxiety involves a fear of being negatively evaluated (Łakuta, 2018), and I predicted it would increase anonymous donations when participants learned their higher frequency of choice, however, this effect was not found. A possible explanation is that, as the study showed participants the proportion of public vs. anonymous donations but not that of no donation, the idea of making a donation in general became more salient, so their best course of action was to donate vs. not donate. No other effects were found for Social Anxiety.

Narcissism – Entitlement/Exploitativeness (NPI-EE) decreased the proportion of Anonymous donations. This result was expected as anonymous donations had no direct benefit for the participant. Participants with higher NPI EE score may feel more entitled to a benefit for any action they perform, something they would not receive from an Anonymous donation. To continue, a higher NPI EE score showed a tendency to decrease the probability of making a donation. These results do not provide evidence of NPI EE increasing prosocial behavior, a main question of this study.

Finally, receiving Social Information predicted a decrease in donation amount. A possible explanation for this result is that, as Social Information influenced the probability of making a donation by its interaction with Social Anxiety, when participants changed their decisions from not donating to making a donation, that donation might have been of a relatively lower magnitude, thus reducing the average.

In conclusion, I found that self-serving traits, such as approval seeking and narcissism, shifted the balance between public, anonymous, and no donations, and that those relationships depended on information about other's decisions. While Simulated Compassion and Social Anxiety were shown to promote prosocial behavior, no evidence was found for Narcissism – which, on the contrary, tended to reduce it. To my knowledge, this is the first study to investigate whether the Submissive (Simulated) Compassion Scale can predict real-life prosocial behavior. The main limitations to the present study include the large proportion of female participants and the limited dissemination of Public donations.

Chapter 2: Self-Conscious Affects

2.1 Background & Significance

Guilt, often referred to as "social glue", has been frequently investigated as an emotion that might mediate charitable giving. However, experimental methods usually involve hypothetical scenarios, which greatly reduces ecological validity, and when monetary transactions are real, they are often based on endowments, which introduces confounds in terms of well-established heuristics (Weaver & Frederick, 2012). Finally, research on guilt and charitable donations usually construes the charity as the object of guilty emotions (e.g., Ty, Mitchell, & Finger, 2017).

In this chapter, I will first review the literature on guilt and prosocial behavior. I will then report a study that employed a real-life guilt-inducing interaction between participants, without experimental deception, to evaluate the influence of guilt on subsequent, and independent, real-life, decisions to donate time & effort in exchange for monetary charitable donations.

2.1.1 Guilt & Prosocial Behavior

Guilt is a self-blaming emotion that involves negative feelings about having broken a social or moral norm (Tangney, 1990; Tangney et al., 1996; Leith & Baumeister, 1998). It is argued that guilt involves a focus on the behavior; the individual feels they have done something

wrong, but that behavior does not define who they are (Tangney, 1990). Moreover, feelings of guilt might motivate apologizing and repairing (Tangney, 1990).

2.1.1.1 The influence of guilt on prosocial behavior

The influence of guilt on prosocial behavior seems to be different for people that have a general tendency to be prosocial in their lives (prosocials), from those that tend to be more selfish or individualistic (proselfs). Ketelaar & Tung Au (2003) had participants play a repeated social bargaining game before and after a guilt manipulation. The manipulation consisted of recalling a recent experience in which they felt guilty, ashamed, or self-blaming. They found that guilt increased cooperative responses most in participants that had previously shown less cooperation. They concluded that guilt serves as meaningful information only to those that had violated a social norm. Nelissen, Dijker, & deVries (2007) carried out a similar experiment in which participants were asked to describe a situation in which they felt very guilty and then played a give-some dilemma game with tickets for a lottery the experimenters would hold at the end of the study. They found that guilt induction increased cooperation for proselfs but not for prosocials. De Hooge, Zeelenberg, & Breugelmans (2007) carried out two analogous experiments in which they had both a guilt and a shame condition induced by recall. Results showed that proselfs in the guilt condition cooperated more than proselfs in the control condition. However, proselfs in the shame condition did not contribute more than proselfs in the control condition. For prosocials there was no effect of guilt or shame. These results support the view of guilt as motivating prosocial behavior, but not shame.

The effect that feeling guilt has on how a person acts in situations that are related to the emotion-eliciting event are considered endogenous influences; while the effect it has on actions that are unrelated to the emotion-eliciting event are considered exogenous. Ketelaar & Tung Au (2003) carried out a second experiment in which they had participants play one round of a social bargaining game and then rate different emotions, including guilt. Participants knew they would be playing a second round of the game with the same participant a week later. Participants were separated into two categories for analysis, specifically, presence or absence of guilty feelings after the first round. Eleven of the twelve participants that made a selfish decision in the first round and felt guilty, made a generous decision in round two. Seven of the nine participants that made a selfish decision in the first round and felt no guilt, made a selfish decision in round two. Based on these results, endogenous influences of guilt also seem to motivate prosocial behavior.

2.1.1.2 Self-conscious Traits: Proneness, Externalization and Detachment

The self-conscious traits of guilt proneness, shame proneness, externalization and detachment play a role in the interpersonal and intrapersonal processes that follow an individual's transgression (Tangney, 1990; Leith & Baumeister, 1998; Uji, Kitamura, & Nagata, 2011). Guilt and shame proneness refer to the predisposition to those emotions after committing a transgression (Cohen, Panter, & Turan, 2012). Externalization involves ascribing the responsibility to other people or aspects of the situation, while detachment implies a low concern or emotional investment in the situation and its outcome (Tangney, 1990). While guilt and shame are affective responses, externalization and detachment seem

to rely more on cognition and can be reactions against painful self-evaluations (Leith & Baumeister, 1998; Uji, Kitamura, & Nagata, 2011).

Guilt proneness and shame proneness have been differentially linked to empathy. While guilt proneness has been linked to the cognitive dimension of perspective taking (Leith & Baumeister, 1998; Cohen et al., 2011), shame proneness has been related to the affective dimension of personal distress (Leith & Baumeister, 1998). Specifically, Leith & Baumeister (1998) did not find guilt proneness to directly predict feelings of guilt; instead, guilt proneness predicted perspective taking, and perspective taking predicted situational guilt. However, perspective taking tended to be absent when feelings of shame were reported. Moreover, situational guilt after an interpersonal conflict seemed to benefit interpersonal relationships, while shame proneness was associated with a deterioration in relationships.

Finally, Cohen, Panter, & Turan (2012) found that guilt proneness predicted the likelihood that a person would behave unethically when choosing between moral and selfish actions. They noted that guilt-prone individuals anticipate feeling guilty about wrongdoings, even if they are private.

2.1.1.3 Guilt, Shame, & Social Conformity

Abell & Gecas (1997) argued that both guilt and shame help individuals to align with group norms of morality and competence. Whereas guilt seems to derive from a commitment to social relationships and its norms; shame appears to be more related to finding oneself to be deficient or incompetent relative to established norms, being concerned about social disapproval, and involving a threat to one's ethical identity (Abell & Gecas, 1997; Harris, 2019).

Silfver (2008) investigated how the relative importance of ten values (e.g., power, benevolence, conformity) related to shame proneness, guilt proneness, and empathy in a sample of adolescents and a sample of military conscripts. Guilt proneness, in both samples, was positively related to valuing conformity. Moreover, a study by Treeby & Bruno (2012) examined shame proneness, guilt proneness, and alcohol use. It found a positive relationship between shame proneness and a motivation to drink due to conformity. In addition, Jiang, Bong, & Kim (2015) looked into the relationship between conformity and several constructs in a population of Korean adolescent students. They found that level of conformity related positively with guilt towards parents when not reaching parental expectations.

2.1.1.4 Guilt & Donations to Charities

Previous studies have found a positive relationship between feelings of guilt and charity donations. Hibbert et al. (2007) performed a study in which participants were presented with a medium guilt-evoking advertisement of a real charity and had to respond to a series of questionnaires. The sample only included individuals that had made a monetary donation to a charity in the past year. They found that the level of guilt a participant felt was positive related to donation intentions. Intention referred to a self-report of how likely the participant was to donate after seeing the advertisement. It is worth mentioning that, although the authors describe the emotion as guilt and used a previously validated scale to measure level of guilt, feeling "ashamed" is one of the items included in the scale.

In Polman & Ruttan (2012) participants were induced guilt (or other emotions) by recalling an experience. Next, they completed a set of surveys unrelated to the study. At the end, they were endowed with 20 cents and asked how much they wanted to donate to cancer research

(and how much they believed others should donate). The cents they did not donate could be exchanged for as many lottery tickets as cents they had left; the lottery had a price of $200. Participants in the guilt, compared to the neutral condition, donated more cents.

Greening et al. (2014) used an fMRI task in which participants assigned intentional and unintentional monetary gains or losses to themselves or a (fictitious) charity. Participants were shown a positive or negative monetary amount and asked to assign it to either themselves or a charity, knowing that one-third of the time the positive amount would become negative and vice versa. Results showed that participants selected the loss trials for themselves slightly more often than for the charity, and selected the gain trials for the charity slightly more often than for themselves. Interestingly, the study found that guilt proneness was positively correlated to the frequency of gains allocated to charities.

Ty, Mitchell, & Finger (2017)'s fMRI study employed a task in which participants view a short description of someone in need from a (fictitious) charity and were then showed a negatively-valenced image and asked whether they wanted to donate certain amount of money to the person in need. After a decision to donate, participants were presented with a neutral/positive feedback screen. After a decision not to donate, participants were presented with guilt inducing feedback that showed adverse consequences for the person in need. In both cases, participants were then given a new chance to donate (usually half the amount of money shown in the first chance). They found that mean ratings of guilt decreased when a not-help decision was followed by a help decision, and when a help decision was followed by another help decision. On the contrary, guilt increased when help was followed by not-help, and when not-help was followed by not-help.

2.1.1.5 Summary

Endogenous and exogenous guilt seem to increase cooperation in social interactions (e.g., Ketelaar & Tung Au, 2003), as well as the probability of donating (e.g., Hibbert et al., 2007), and the donation amount (e.g., Polman & Ruttan, 2012). In the next section, I discuss how the present study fills a gap in the current literature, with respect to ecological validity, as well as the transfer of guilt in one domain to prosocial decisions in another.

2.1.2 A Study on the Influence of Guilt on Charitable Donations

The existing literature has used three main strategies to elicit guilt: 1) autobiographical recall (Tangney et al., 1996; Ketelaar & Tung Au, 2003; Schmader & Lickel, 2006; De Hooge, Zeelenberg, & Breugelmans, 2007; Nelissen, Dijker, & deVries, 2007; De Hooge, Breugelmans, & Zeelenberg, 2008; Polman & Ruttan, 2012), 2) scripts of hypothetical scenarios (De Hooge, Breugelmans, & Zeelenberg, 2008; Nelissen, 2014); and 3) causing a wrong-doing, or making the participant believe they have caused one, during the experiment (De Hooge, Breugelmans, & Zeelenberg, 2008; Nelissen, 2014; Yu et al., 2014; Ty, Mitchell, & Finger, 2017).

An issue with the autobiographical recall experiments is that most of them rely on the participant's understanding of the meaning of guilt vs. shame, and on their ability to recall the experiences accurately. In the case of hypothetical scenarios, participants might not feel or respond the same way as in situations where a real transgression is committed. In the case of using deception to make participants believe they have caused a wrongdoing, other emotions might be elicited if participants suspect the feedback they are receiving does not

coincide with their perception of how they are doing. Inducing self-blaming emotions with an interpersonal task and real feedback, particularly in research about charity donations, is an important contribution of the present study.

In addition to memory constraints and lack of realism in guilt induction, many studies fail to use real-life consequences when measuring prosocial behavior (e.g., De Hooge, Zeelenberg, & Breugelmans (2007) did not use real money). Moreover, many of the tasks involve strategy, risk, or trust – as the results depended on other participant's decisions as well (e.g., Nelissen, Dijker, & deVries, 2007) –, which can also influence decisions. Regarding studies involving donations, they usually employ an endowment that participants can keep if they decide not to donate (e.g., Polman & Ruttan, 2012). In the present study, I use a task in which participants can gain money for charities by giving their time and effort. A decision not to donate does not involve a higher monetary gain to the participant.

In the current study, guilt induction involves dyads of participants randomly assigned the role of Advisor or Decider in a task in which the Decider had to accept the Advisor's judgment regarding the number of a set of briefly flashed dots, and suffered negative consequences, in terms of time and effort, if that judgment was wrong. The difficulty of the task, Hard vs. Easy, determined the number of incorrect judgments provided by the Advisor. Following the dot judgment task, both participants were given the opportunity to donate time and effort to a real charity. In a Social Information condition, Advisors were informed of the performance of other Advisors on the dot task before proceeding to the charitable donation phase. I predicted that Advisors in the Hard, relative to Easy, condition would be more likely to donate during the charity phase, because of a general sense of guilt. I further predicted that

this effect would be greatly reduced given information about the performance by other participants in the Social Information condition.

2.2 Methods

2.2.1 Participants

One hundred and eighty participants (76 female, mean age 37.92 ± 11.88) completed the study on Prolific (www.prolific.com). Participants were paid $5 for thirty minutes of participation. Participants had the option to complete additional tasks in exchange for charity donations without direct monetary compensation for themselves. All participants gave informed consent and the Institutional Review Board of the University of California, Irvine approved the study. Participants were assigned to one of three different conditions of the emotion eliciting task (Easy, Hard, or Hard & Social Information), were divided into pairs, and then were assigned a role within those pairs (Advisor vs. Decider), for a total of six possible different categorizations (e.g., advisor in the easy group). Data from thirty participants was collected for each category. The sample size was determined by the number of participants needed to detect a 0.5 correlation with 80% power (Bujang & Baharum, 2016).

2.2.2 Tasks

Dots Task This interpersonal task was based on Yu et al. (2014) and Zhu et al. (2019). Participants were divided into pairs and randomly assigned the role of Advisor or Decider. On each round, both participants saw a white screen with black dots for 1.5 seconds and were asked to estimate how many dots were presented. They knew that in half of the rounds the Decider would have to submit a decision on their own (rounds without advice); while on the other half, the Advisor would send an advice to the Decider about what they believed the correct decision to be, and the Decider would have to follow it (rounds with advice). In the Easy condition, participants had to select whether the number of dots was below or above 20. In the Hard and Hard & Social Information conditions, participants had to select if the number of dots was exactly 17, 19, 21, or 23. The number of dots were set across rounds with and without advice in such a way that the difficulty of the task was analogous to both participants. While incorrect responses by the Decider had no effect, incorrect responses by the Advisor had a negative effect on the Decider. For every incorrect response in a round with advice, the Deciders would have to complete three additional rounds of the task on their own once the original rounds were over. The pair of participants completed 30 rounds together. Once these rounds were over, the Advisor was able to move forward with the following task, but the Decider had to complete additional rounds of the Dots task by themselves equal to 3 times the number of incorrect advice (i.e., anywhere between 0 and 45 extra rounds).

Charity Task Participants were presented with the name, an image, and mission of 20 real charities, one per round. In each round participants had to decide whether they wanted to spend extra time at the end of the study to earn money for the presented charity. They were instructed that to earn money for the charity they would have to correctly complete Slider Tasks (see below). The participants had to enter a numeric response between 0 and 20, referring to the number of Slider Tasks they committed to complete. For each Slider Task, they could earn $1 for the charity. Participants knew only one of the rounds would be randomly selected and they would be asked to complete the number of Slider Tasks they committed to in that round. The charities were sent the money gained by the participants.

Slider Task Participants were presented with five sliders positioned randomly across the screen. All sliders began with their value set at zero (left of the slider), and participants had 30 seconds to move all values to fifty (middle of the slider). If, when the timer, ran out all sliders had a value of fifty, the participant would earn $1 for the selected charity. Participants could complete up to 20 Slider Tasks, with the position of the sliders on the screen being random at the beginning of each task. This task is identical to the one represented in Figure 1.3.

2.2.3 Procedure

Participants began the study by completing the self-report measure (see below). They were then paired with another participant and told which role they were assigned. They first completed 3 practice rounds of the Dots Task on their own and then completed 30 rounds together. Once those rounds finished, participants were told how many incorrect advice they

gave or received, and how many extra rounds the Decider would have to complete. In the Hard & Social Information condition, participants were also shown the statement "In a sample of 30 previous Advisors, the ***average number of incorrect advice was 8.13***." This number corresponded to the average number of incorrect advice given by the Advisors from the Hard condition. On the same page, in all conditions, participants were asked to respond how they would characterize their performance as Advisor (or the Advisor's performance in the case of the Deciders), on a five-point Likert scale ranging from "Very poor" to "Very good". After this, Deciders completed any additional rounds before moving to the Charity Task. Advisors moved directly to the Charity Task. Participants were then thanked and told they would be sent $5 for participating in the study. On the same page, they were told which charity was chosen and the number of Slider Tasks they committed to performing. They were reminded that they would not receive any additional monetary compensation for the time they spent completing the Slider Tasks, but that they would be donating their time and effort to gain money for the charity. If the participant had chosen a positive number of Slider Tasks for the selected charity, they would then proceed to complete them as the last part of the study.

Four pilots (N=48) were conducted to design the Easy and Hard conditions such that the Advisors in the Easy (E) group would give a small number of incorrect advice and thus no consequential emotion would be elicited. In contrast, in the Hard group (H), Advisors would give a considerable number of incorrect advice and would feel guilty because of being responsible for the Decider having to complete additional rounds of the task and making their experiment longer. Prolific participants get paid for each experiment they complete, the more time they spend on a study, the less time they have to spend on other studies to

earn more income. Finally, in the Hard & Social Information (HSI) group it was expected that the guilt-inducing power of incorrect advice would be reduced, since the vast majority of Advisors performed similarly on the task – a form of herd immunity.

The feelings of guilt were expected to be short-lived due to their origin and intensity. Ketelaar & Tung Au (2003) found their self-blame manipulation to only last for the first 10 rounds of their task. Given that each round of their task involved two steps, while the Charity Task from the present study only involved one (i.e., participants saw the charity and input their response in the same screen), the total number of rounds was set to 20.

2.2.4 Self-Report Measures

Test of Self-Conscious Affects – 3 (Tangney et al., 2000) This measure presents participants with a series of scenarios that people are likely to encounter in daily life as well as a series of possible responses a person might have in those scenarios. For each possible response, participants have to rate how likely they are to react or feel that way. The short version that includes eleven negative scenarios was used in the current study, providing scores for guilt proneness, shame proneness, externalization, and detachment. Scores for each subscale can range from 11 to 55.

2.2.5 Statistical Analyses

All statistical analyses were implemented in JASP and MATLAB. Two two-way analyses of variance (ANOVA) were performed to compare the mean difference in proportion of

donations (i.e., number of rounds in the Charity Task in which participants entered a positive donation over the total number of rounds) and average magnitude of donations on rounds with a positive donation (i.e., average number of Slider Tasks across rounds with positive donations) by advisors in the two Hard groups, with and without social information about norm performance, and with participants divided into additional groups based on their performance relative to the norm (below vs. at/above).

Two one-way ANOVAs were performed including all groups, Hard (H), Hard & Social Information (HSI), and Easy (E), assessing the mean proportion and magnitude of donations, respectively, as a function of group. As criterion checks, two additional ANOVAs compared the objective advisor performance and self-rated advisor performance (on a Likert scale converted to values from -2 to 2, where a higher value indicated worse rated performance) between conditions.

Correlation analyses using Pearson's *r* were performed with proportion and magnitude of donations as outcome variables and with objective performance, rated performance, guilt proneness, shame proneness, externalization, and detachment as predictor variables. Guilt proneness was partialed out of shame proneness and vice versa, and externalization was partialed out of detachment and vice versa.

In the case of participants assigned the role of Decider, analogous ANOVAs and correlation analyses were performed. Correlation analysis also included as predictor variable the number of incorrect decisions made by the Deciders in rounds without advice. However, as these are not of primary interest, the results are reported in Supplemental Tables 2.1 and 2.2.

2.3 Results

Of primary interest was whether objective performance on the advising task would modulate the tendency of Advisors to donate in the subsequent charity task, even though the tasks were independent with respect to the Decider, the presumed object of the Advisor's guilt. Moreover, I expected knowledge about the performance of other Advisors to modulate these effects.

First, the ANOVA with Social Information & Performance Relative to the Norm as factors yielded a significant interaction, such that the proportion of donations was greater in advisors that performed worse relative to the norm than in advisors that performed better than the norm when no norm information was provided, while the proportion of donations was *smaller* in advisors that performed worse relative to the norm than in advisors that performed better than the norm when advisors were informed about their performance relative to the norm (see Figure 2.1), $F(1,56)=5.7$, $p<0.03$. Post-hoc comparisons revealed the only significant difference to be between the H and HSI Above the Norm groups ($t(37)=3.5$, $p_{tukey}=0.005$).

The ANOVA comparing the proportion of donations across E, H, and HSI groups [0.72 (0.07), 0.77 (0.06), and 0.52 (0.08), respectively], yielded a significant effect of group [$F(2,87)=3.588$, $p=0.032$]. Again, post-hoc comparisons revealed the only significant difference to be between the H and HSI groups ($t(58)=2.526$, $p_{tukey}=0.035$). A participant's probability of making donations decreased, on average, 25 percentage points in the HSI

group. There were no significant effects involving the mean magnitude of donations in donation rounds.

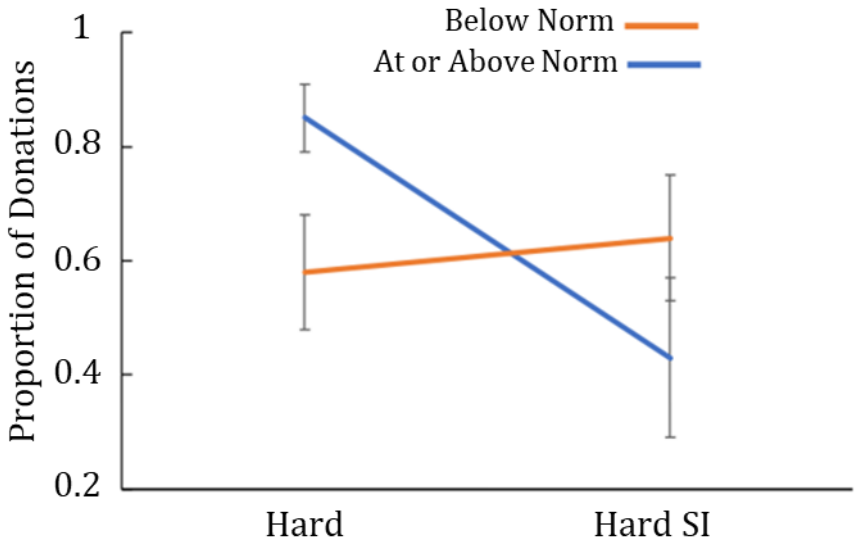

Figure 2.1 Results. Mean proportion of rounds in which Advisors donated, in the absence vs. presence of social information about their performance relative to norm. Error bars = SE

Second, in the Hard group, an increase in the number of incorrect advice predicted an increase in the number of subsequent, independent, donations ($r=0.38$, $p<0.05$). Notably, a trend in the *opposite direction* was observed for the magnitude of donations, which decreased as the number of incorrect advice increased ($r=-0.21$): though this latter correlation did not reach significance ($p=0.29$), the correlation coefficient, going in the opposite direction, was significantly different from that involving the number of donations ($p<0.03$). In the Easy group, the magnitude of donations significantly decreased with a

decrease in self-rated performance (r=-0.45, p<0.03), such that advisors donated less the worse they perceived their own performance. Similar, though non-significant, trends were observed for the objective accuracy of performance, with respect to both the proportion and magnitude of donations – worse performance predicted a reduction in prosocial behavior in the Easy group (see Table 2.1). Finally, in the group receiving social information about their peers' performance (HSI), the only effects close to significance were obtained when dividing the group into Advisors that performed at/above vs. below the social norm: though such divisions yielded too small samples to interpret, correlations in these subgroups are reported in Table 2.1 for completeness.

Table 2.1 Results for Advisor Participants. Correlation coefficients for the influence of objective and self-rated advisor accuracy on the proportion and magnitude of donations by Advisors in the Hard (H), Easy (E), and Hard & Social Information (HSI) groups. HSI Below and HSI Above only include the participants from the HSI group that gave a number of incorrect advice below and at/above average, respectively. *p≤0.10, **p≤0.05.

	Prop. of Donation	**Magn. of Donations**
Number Incorrect Advice		
E (n=30)	-0.13	-0.33
H (n=30)	**0.38****	-0.21
HSI Below (n=12)	0.32	0.44
HSI At/Above (n=18)	-0.14	-0.19
Rated Advisor Performance		
E (n=30)	-0.17	**-0.45****
H (n=30)	0.21	-0.24
HSI Below (n=12)	**0.56***	0.22
HSI At/Above (n=18)	-0.12	-0.42

With respect to trait measures, the proportion of donations increased with Guilt Proneness in the E group (r=0.44, p<0.02). The tendency was also found in the H group (r=0.33,

p=0.083). Inversely, Shame Proneness negatively correlated with the proportion of donations in the E group ($r=-0.38$, $p<0.05$), and the same tendency was found for Externalization ($r=-0.31$, $p=0.098$). Detachment had a positive correlation with the magnitude of donation in the E group ($r=0.41$, $p<0.05$). In the HSI group (all participants), the magnitude of donation increased with Externalization ($r=0.47$, $p=0.02$). See Table 2.2 for correlations dividing the HSI group into Advisors that performed at/above vs. below the social norm.

Additional analyses confirmed that objective performance differed as expected across groups ($F(2,87)=17.846$, $p<0.001$, $\eta^2=0.291$). The mean number of incorrect advice was 4.53 (0.41) in the Easy group, 8.13 (0.53) in the Hard group, and 7.97 (0.49) in the HSI group. Post-hoc comparisons revealed significant differences between the E and the H groups ($t(58)=-5.292$, $p_{tukey}<0.001$) and the E and HSI groups ($t(58)=-5.047$, $p_{tukey}<0.001$) but not between the H and HSI groups ($t(58)=0.245$, $p=0.967$), showing only the effect of difficulty. The self-rated performance of the advisor was 0.00 (0.2), 0.73 (0.2), 0.00 (0.3), and 0.89 (0.2) for the E, H, HSI below social norm, and HSI above social norm groups, respectively (note that a higher average indicates worse performance) ($F(3,86)=4.289$, $p<0.01$, $\eta^2=0.13$). Post-hoc comparisons revealed a significant difference between the E and H ($t(58)=-2.693$, $p_{tukey}<0.05$) and E and HSI above social norm ($t(46)=-2.827$, $p_{tukey}<0.03$) groups. These results provide evidence of the Advisors paying attention to and taking into consideration the social information manipulation from the experimental design.

Table 2.2 Correlation Coefficients for Trait Measures and Donation Decisions. HSI Below and HSI Above only include the participants from the HSI group that gave a number of incorrect advice below and at/above social norm, respectively. *p≤0.10, **p≤0.05

	Prop. of Donations	**Magn of Donations**
Guilt Proneness		
E (n=30)	**0.44****	0.12
H (n=30)	**0.33***	0.13
HSI Below (n=12)	0.50	0.25
HSI At/Above (n=18)	0.18	-0.04
Shame Proneness		
E (n=30)	**-0.38****	-0.02
H (n=30)	-0.04	0.07
HSI Below (n=12)	-0.09	0.10
HSI At/Above (n=18)	-0.32	0.17
Externalization		
E (n=30)	**-0.31***	-0.16
H (n=30)	-0.05	0.09
HSI Below (n=12)	0.22	0.47
HSI At/Above (n=18)	-0.05	0.41
Detachment		
E (n=30)	0.20	**0.41****
H (n=30)	-0.22	-0.23
HSI Below (n=12)	0.21	0.03
HSI At/Above (n=18)	0.17	-0.10

2.4 Discussion

This study looked at the relationship between self-conscious affects, social information, and donations to charities. Participants completed one of three conditions of an interpersonal task in which the decisions of one of the members of the pair (Advisor) could negatively

impact the other person (Decider) but not themselves – that is, having to spend more time completing the study for the same monetary compensation. The Easy (E) condition had a low probability of error so that no negative self-conscious emotion would be elicited. The Hard (H) condition was designed so that the probability of error was higher, to elicit guilt due to the emphasis on their behavior being the reason the Decider had to spend more time on the experiment. Lastly, the Hard & Social Information (HSI) condition had the same level of difficulty as the H condition, but the emphasis was shifted from the Advisors' behavior having a negative effect, to how the Advisors' performance compared to that of a previous group of Advisors. After the interpersonal task, participants had to make decisions about donations to charities. I employed a task in which participants did not donate money directly, instead, the decision was whether to donate their time and effort to gain money for the charity. I looked at how each condition of the interpersonal task distinctly impacted Advisors' decisions in the Charity Task.

A significant interaction between Social Information and Performance Relative to the Norm was found for the two groups that completed a hard version of the dots task. Participants that performed worse than the norm donated more when they did not know how other Advisors performed than when that information was provided. This can be interpreted in two different ways. Participants that performed worse than the norm and were then informed about the norm might have changed their reference point about what a poor performance consisted of. The distance between their number of incorrect responses and what they thought was expected shrunk, meaning they caused the Decider less harm than originally thought and felt less guilty. Another interpretation could be that, when learning about the norm, feelings of guilt were exceeded by feelings of shame, which has been related

to finding oneself to be incompetent relative to established norms (Abell & Gecas, 1997). In a study by Ibanez & Roussel (2021), results pointed towards feelings of shame reducing average donations towards a non-governmental organization. In the present study, feelings of shame might have decreased the proportion of donations.

The number of incorrect advice had a moderate positive correlation with the proportion of donations in the Hard condition. This point to feelings of guilt increasing the probability that a person would donate. This result gives further support to Hibbert et al. (2007), where the level of guilt a participant felt was positive related to a participant's self-report of how likely they were to donate. While guilt increased the proportion of donations, there was a tendency to reduce the magnitude of donations, which provides evidence to guilt not necessarily influencing every aspect of prosocial behavior equally. This result differs from Polman & Ruttan (2012), where participants in a guilt, compared to a neutral condition, donated more. Notably, the main aim of their study as well as their task was different from the present study.

Guilt proneness was positively associated with proportion of donations in the E and H groups. Guilt proneness has been linked to perspective taking (Leith & Baumeister, 1998), hence, people higher in guilt proneness might be better able to focus on the need of the populations served by the charities and increase the probability of donation. Regarding shame proneness and externalization, they had a negative relationship with proportion of donations in the Easy group. The fact that these relationships were mainly seen in the Easy condition might relate to the fact that this was the only group in which there was no emotion manipulations, so a direct influence of shame proneness and externalization can be better

assessed. Shame proneness has been significantly correlated with the personal distress dimension of empathy (Leith & Baumeister, 1998). When looking at the charities, higher shame prone people might have a tendency to change the focus from the other to the self, thus reducing the probability of making a donation. Moreover, externalization involves not ascribing responsibility to oneself, so a lower probability of donation is expected. Surprisingly, there was a positive correlation between detachment and amount donated in the Easy group.

The present study investigated the effects of exogenous guilt, given that the decision to donate to charity was not related to the emotion eliciting event. Although guilt has been generally considered as a negative emotion that leads to positive interpersonal consequences (e.g., Tangney, 1990), a study by de Hooge (2012) suggested that the main goal of guilt reparations might not be the welfare of the person damaged but getting rid of the feeling itself. This might be one of the reasons why exogenous guilt seems to have the same effects as endogenous guilt; if the welfare of the person damage is not the main concern, then the reparative behavior does not need to go towards them. Results from the current study of guilt increasing the probability of making a donation can be interpreted in two different ways. As participants were not able to direct reparation towards the individual, they may have chosen to act prosocially towards the charities to compensate. Another possibility is that, as participants were not able to recompense the Decider, they decided to punish themselves by spending more time on the experiment, as donations entailed extra time and effort. Nelissen & Zeelenberg (2009) showed that self-punishment can occur in participants that are not able to compensate the victim of their transgression.

One of the limitations of the present study is that I did not include a measure of the emotions the participants were feeling after each task. Instead, participants were assumed to be likely to respond with feelings of guilt depending on which condition and role they were placed in, based on the existing literature. Notably, significant results pointed towards the emotion manipulations working as expected. Importantly, because I did not measure feelings directly, not everyone in the Hard conditions might have felt guilt, which would imply the effects of these emotions might be underestimated in the current analysis. Moreover, it is important to mention that the sample size was only able to detect moderate-to-strong and larger correlations; for this reason, there are many correlations in the results that had moderate size coefficients but do not reach significance.

In conclusion, despite some limitations, this work adds relevant empirical evidence to the literature on self-conscious emotions, social conformity, and charity donations.

Supplemental Table 2.1 Results for Decider Participants: ANOVAs and Post Hoc Tests.

	F or t	p	η^2
Proportion of Donations			
(With vs. Without Social Information) x	F(1,56)=1.119	0.295	0.02
(Below vs. At/Above Social Norm)	F(1,56)=0.11	0.741	0.00
	F(1,56)= 0.026	0.873	0.00
(Easy vs. Hard vs. Hard & Social Information)	F(2,87)=1.528	0.223	0.03
Magnitude of Donations			
(With vs. Without Social Information) x	F(1,49)=2.514	0.119	0.049
(Below vs. At/Above Social Norm)	F(1,49)=0.048	0.828	0.00
	F(1,49)=0.17	0.682	0.00
(Easy vs. Hard vs. Hard & Social Information)	F(2,77)=1.688	0.192	0.04
Objective Decider Performance			
(Easy vs. Hard vs. Hard & Social Information)	F(2,87)=32.592	<0.001	0.43
Post Hoc Test Easy vs. Hard	t(58)=-7.652	<0.001	
Easy vs. HSI	t(58)=-6.056	<0.001	
Hard vs. HSI	t(58)=1.597	0.253	
Rated Advisor Performance			
(Easy vs. Hard vs. HSI below vs. HSI above)	F(3,86)=3.494	0.019	0.11
Post Hoc Test HSI below vs. HSI above	t(28)=-2.385	0.088	

Supplemental Table 2.2 Results for Decider Participants. HSI Below and HSI At/Above only include the participants from the HSI group that gave a number of incorrect advice below and above social norm, respectively. Number Incorrect Decisions are the incorrect decisions in the rounds without advice. *p≤0.10, **p≤0.05, ***p≤0.01.

	Prop. of Donations	Magn. of Donations
Number Incorrect Decisions		
E (n=30)	-0.24	**-0.35***
H (n=30)	-0.09	-0.19
HSI Below (n=12)	0.15	-0.45
HSI At/Above (n=18)	0.18	-0.15
Number Incorrect Advice		
E (n=30)	0.19	-0.08
H (n=30)	0.06	-0.03
HSI Below (n=12)	-0.34	-0.03
HSI At/Above (n=18)	-0.38	0.26
Rated Advisor Performance		
E (n=30)	-0.01	0.14
H (n=30)	0.23	-0.02
HC Below (n=12)	**-0.79*****	-0.54
HC At/Above (n=18)	-0.02	0.21
Guilt Proneness		
E (n=30)	0.19	**0.34***
H (n=30)	0.27	0.07
HSI Below (n=12)	**0.64****	**0.67****
HSI At/Above (n=18)	**0.42***	0.00
Shame Proneness		
E (n=30)	0.22	0.19
H (n=30)	-0.03	0.13
HSI Below (n=12)	-0.40	**-0.63***
HSI At/Above (n=18)	**-0.46***	-0.42
Externalization		
E (n=30)	-0.28	-0.05
H (n=30)	0.10	-0.25
HSI Below (n=12)	-0.09	-0.02
HSI At/Above (n=18)	-0.36	0.22
Detachment		
E (n=30)	0.19	-0.13
H (n=30)	**-0.34***	0.32
HSI Below (n=12)	**-0.61****	**-0.72****
HSI At/Above (n=18)	0.26	-0.34

Chapter 3: Control

3.1 Background & Significance

The degree to which voluntary actions yield distinct consequences has recently been identified as an essential aspect of dynamic reward maximization (see Liljeholm, 2021 for review). As an illustration, imagine that you are choosing between 3 vending machines, each located in a different corner of campus, at the same, substantial, distance from your current location. For one machine, the selection of one of two options is computerized, such that you cannot decide yourself if you are getting soft drink A or B; for another, you can select between two buttons yourself, but both selections yield soft drink A. Finally, for the third machine, buttons 1 and 2 yield soft drinks A and B, respectively, and it is up to you to select a button. Intriguingly, according to conventional theories of choice, as long as you value drinks A and B equally at the time of deciding between vending machines, all are equally good options – in contrast, a plethora of research has demonstrated that animals, including humans, prefer to freely choose between options (Bown, Read, & Summers, 2003). Moreover, recent research (Mistry & Liljeholm, 2016; Liljeholm et al., 2018; Norton & Liljeholm, 2020; Liljeholm, 2022) suggests that the third vending machine, which allows you to choose freely *and* produce different outcomes depending on your decision, is preferred over the other two. In other words, people prefer environments with greater differences between outcome distributions associated with freely chosen actions.

Liljeholm and colleagues commonly formalize instrumental divergence as the Jensen-Shannon (JS) divergence of outcome probability distributions associated with alternative, and freely chosen, actions. If 'P_1' and 'P_2' are the outcome probability distributions of two available actions, 'O' is the set of possible outcomes, 'o' is one possible outcome, and 'P(o)' is the probability of that outcome, then

$$ID = \frac{1}{2}\sum_{o \in O} log\left(\frac{P_1(o)}{P_*(o)}\right) P_1(o) + \frac{1}{2}\sum_{o \in O} log\left(\frac{P_2(o)}{P_*(o)}\right) P_2(o),$$

where $P_*(o) = \frac{1}{2}(P_1 + P_2)$. Note that, although JS divergence can be computed over probability distributions of any type of random variables, it is not instrumental unless referring to the objective outcomes of freely chosen actions (Liljeholm, 2022).

At the neural level, a signal in the supramarginal gyrus of the inferior parietal lobule has been shown to scale with trial-by-trial changes in instrumental divergence (Liljeholm et al., 2013), as well as with high- vs zero-divergence instrumental schedules, that yield significantly different levels of goal-directedness (Liljeholm et al., 2013). Furthermore, Norton & Liljeholm (2020) used a task in which participants chose between environments with different instrumental divergence, expected monetary payoffs, and free vs. forced choice. They found that a measure of expected value that included instrumental divergence as a reward surrogate performed better than a conventional utility model, and that activity in the ventromedial prefrontal cortex scaled with this divergence-based expected utility.

Notably, all previous studies assessing the intrinsic utility of instrumental divergence have focused on decisions that impact the agent themselves, yet a great number of everyday decisions are motivated by their impact on others. The present study sought to investigate whether the preference for high instrumental divergence is also present when individuals are making decisions that only have a direct consequence for charitable organizations.

3.2 Methods

3.2.1 Participants

One hundred (32 female, mean age 36.76 ± 11.57) completed the study on Prolific (www.prolific.com). Participants were paid $10 for one hour of participation and could earn up to an additional $36 for themselves, charity, or both. All participants gave informed consent and the Institutional Review Board of the University of California, Irvine approved the study. The sample size was determined by Liljeholm (2022) and considering a power of 0.80 to be sufficient.

3.2.2 Task

The task is based on Liljeholm (2022). Participants were told to assume the role of a gambler playing slot machines in a casino with the objective of gaining as much money as possible in each round. The earnings of each round were either for themselves or for a charity they would later choose from a large pool. At the beginning of each round, participants were told

who would receive the earnings of that round and had to choose one of two rooms to gamble in. Each room had only two slot machines available, and once a participant chose a room, they could only gamble in that room for the 4 consecutive trials that made up the round. Rooms were also classified as self-play or auto-play. In self-play rooms, participants could choose either of the two slots machines in each of the four trials. In auto-play rooms, the slot machine that was chosen in each trial was predetermined; slots machines were alternated between trials, so each machine was selected two times. Slot machines produced three different color tokens (red, blue, or green) with different probabilities. Slot machines were represented as pie charts, with each colored slice within the pie chart representing the probability that the machine will produce the respective color. The amount of money that each color token was worth changed from round to round and was only told to the participants after they had already chosen the room to gamble in. Participants knew that 3 of a total of 48 rounds would be randomly selected and either themselves, the charity, or both, would be paid the money earned across those rounds.

The difference between outcome distributions associated with the slot machines in a room was formalized as the JS divergence of the room, with 4 such differences (0.00, 0.04, 0.15, and 0.20) yielding 6 unique combinations between the two rooms presented at the beginning of a round (0.00 vs. 0.04, 0.00 vs. 0.15, 0.00 vs. 0.20, 0.04 vs. 0.15, 0.04 vs. 0.20, and 0.15 vs. 0.20). JS divergence conditions were combined with 4 self- vs. auto-play combinations (both rooms self-play, both rooms auto-play, greater divergence room self-play, and lesser divergence room self-play). These room combinations were each presented one time as earning for the participant and one time as earning for the charity, for a total of 48 rounds. The order in which each combination was presented was random across all 48 rounds. Each

round had 4 gambling trials for a total of 192 gambling trials. Possible monetary values for the color token ranged from -$2 to $3. See Figure 3.1 and Figure 3.2 for screenshots of the task.

Figure 3.1 Screenshot of the Beginning of a Round. Participants were told who would receive the earnings of the round (charity in the screenshot) and shown two rooms (in the screenshot, 0.00 JS divergence and auto-play on the left, 0.20 JS divergence and self-play on the right). Note that both pie charts look the same in the case of 0.00 JS divergence. The room on the left has zero instrumental divergence, while the one on the right has high instrumental divergence. Participants looked at this information and chose one of the rooms to gamble in the next 4 consecutive trials.

Figure 3.2 Screenshot of a Gambling Trial. Participants were shown the two slot machines available in the room they selected and the monetary values of each color token for that round. The screenshot shows the case of an auto-play room, in which participants were told which slot machine to select by surrounding it with a square and showing a prompt underneath (the one on the left in the screenshot) – in self-play rooms none of the slot machines were surrounded by a square and a prompt appear under both slot machines. After selecting a slot machine, the outcome of the trial was shown (e.g., $1), before proceeding to the next trial or round.

The primary measure of the study was the decision at the beginning of each round between gambling rooms. Rooms differed in terms of JS divergence as well as instrumental control. Self-play rooms with higher level of JS divergence had higher instrumental divergence, but auto-play rooms, not matter the level of JS divergence, had no instrumental control because slot machines were not freely chosen within the round, so they had no instrumental divergence. Higher JS divergence in the task was associated with higher outcome diversity. Research has shown the preference of people for outcome diversity, even willing to select more highly diversified options that had lower expected utility (Ayal & Zakay, 2009). Auto-play rooms with higher JS divergence conserved the outcome diversity of self-play rooms with higher JS divergence but eliminated instrumental divergence. Including auto-play rounds thus helped eliminate outcome diversity as a confounding variable. To also eliminate expected monetary payoffs and outcome entropy (uncertainty about what outcome will occur) as the main reason behind participants' decisions, and even though participants did not know the token monetary values until after selecting a room, the two variables were programmed so that they would be relatively constant across rooms, or in the case of expected monetary payoff, slightly biased against the influence of JS divergence. See Table 3.1 for the programmed values.

Table 3.1 Liljeholm (2022) Table 1. Mean Shannon entropy and expected monetary payoffs at each divergence level.

Room Divergence	**0.00**	**0.04**	**0.15**	**0.20**
Mean Entropy	0.74	0.71	0.71	0.71
Self-play payoff	$0.39	$0.36	$0.34	$0.38
Auto-play payoff	$0.39	$0.37	$0.34	$0.37

3.2.3 Procedure

Participants first entered their demographic information. They then read the instructions and completed the 48 rounds of the task. After that, they filled out the self-report measures described in the next session. Finally, they selected a charity from a pool of 60 charities.

3.2.4 Self-Report Measures

Social Anxiety Questionnaire (Łakuta, 2018) (SAQ) This scale measures social anxiety defined as "a marked and persistent fear of negative evaluation in social situations". The scale consists of 10 statements and participants responded on a five-point scale ranging from "Strongly Disagree" to "Strongly Agree". Scores can range from 10 to 50.

Test of Self-Conscious Affects – 3 (Tangney et al., 2000) This measure presents participants with a series of scenarios that people are likely to encounter in daily life as well as a series of possible responses a person might have in those scenarios. For each possible response, participants have to rate how likely they are to react or feel that way. The short version that includes eleven negative scenarios was used in the current study, providing scores for guilt proneness, shame proneness, externalization, and detachment. Scores for each subscale can range from 11 to 55.

3.2.5 Statistical Analyses

The following tests were performed: 1) Two-way repeated measures ANOVA for choice proportions on 0.20 JS divergence rooms with recipient (participant vs. charity) and play

condition (self vs. auto) as independent variables. Only conditions with high outcome divergence and self-play reflect instrumental control. 2) Two-way repeated measures ANOVA for choice proportions on 0.00 JS divergence rooms with recipient (participant vs. charity) and play condition (self vs. auto) as independent variables. These analyses control for mode of play. All statistical analyses were implemented in JASP (https://jasp-stats.org/).

3.3 Results

Consistent with the notion that the preference for agency differs across earning for oneself vs. others, there was a significant interaction between recipient (participant vs. charity) and play condition (self vs. auto) when the rooms had high outcome divergence ($F(1,99)=21.11$, $p<0.001$, $\eta^2=0.049$). Post hoc tests revealed that choice proportions for participant-self-play were significantly higher than participant-auto-play, as well as charity-self-play and charity auto-play (greatest $p=0.002$). Moreover, participant-auto-play was significantly lower than charity-auto-play ($p_{holm}=0.003$) and charity-self-play ($p_{holm}=0.001$), again suggesting a preference for free choice when choosing for oneself more than when choosing for others. Charity-self-play and charity-auto-play were not significantly different. Proportions are shown in Figure 3.3.

There was also a significant interaction between recipient (participant vs. charity) and play condition (self vs. auto) in rooms with zero outcome divergence ($F(1,99)=4.43$, $p=0.038$, $\eta^2=0.016$). Post hoc tests revealed the only significant difference to be between participant-

self-play and participant-auto-play (p_{holm}=0.041), once again suggesting a preference for free choice when choosing for oneself but not when choosing for others. Proportions are shown in Figure 3.3.

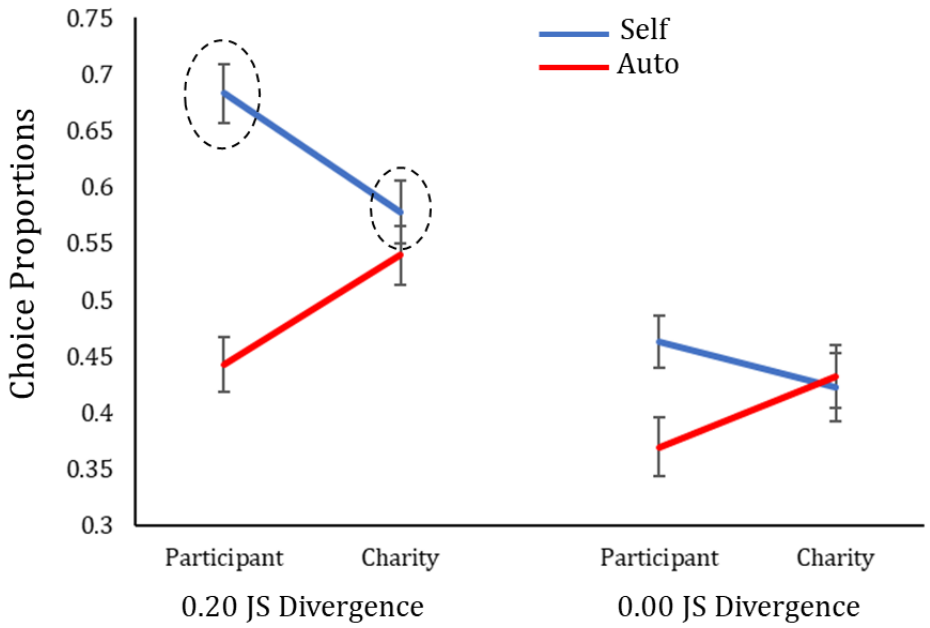

Figure 3.3 Results. Proportion of selecting a gambling room given JS divergence (0.20 vs. 0.00), self- (blue) vs. auto-play (red), when gambling for charity vs. oneself. Note that instrumental divergence is only high given both self-play and high (0.20) JS divergence (indicated by dashed circles). Error bars = SE.

3.4 Discussion

The current study researched the role of instrumental divergence in decision making that has an effect on charities as opposed to oneself. Adapting the gambling task from Liljeholm (2022) to include both rounds where potential earnings were for the participant and rounds

for charities, the preference for environments with high instrumental divergence was found weakened when decisions did not have an influence on oneself.

3.4.1 Free Choice & Outcome Divergence

When earnings were for the participant, the proportion of choices was higher in self- vs. auto-play rooms in both high and low outcome divergence rooms. Environments with free choice can have a higher probability of allowing people to avoid potential negative scenarios. For example, looking at the slot machines available in the right room of Figure 1, one of the machines does not yield blue tokens. If a person is in this room and learns that blue tokens result in -$2, when the room is self-play the person is able to completely avoid blue tokens by selecting the other slot machine, however, if the room is auto-play, the person cannot completely avoid the blue tokens. This preference for free choice was not seen when earnings were for charity – the proportion of choices did not differ between self- and auto-play rooms with high and low outcome divergence. Although free choice is generally preferred to forced choice, it requires additional cognitive processes (e.g., Naefgen, Dambacher, & Janczyk, 2018). This additional cost has a higher probability of being rewarded in situations in which the outcomes can generate a direct benefit for the person making the effort, than when benefits would go to charity.

When earnings were for charity, participants tended to prefer rooms with higher outcome divergence, regardless of self- vs. auto-play conditions. These decisions may have been driven by a diversity heuristic, in which the higher the perceived diversity the lower the risk (Ayal & Zakay, 2009). The perceived diversity of outcomes within rooms could be quickly

evaluated due to the slot machines being represented as colored pie charts that yield colored tokens, and participants might have based their decisions on a heuristic that biased their decisions towards higher perceived diversity, rather than, again, more effortful computations.

Whereas free choice and outcome diversity influence choice in their own right, I am primarily interested in the influence of the intersection – i.e., instrumental divergence – on prosocial behavior. In the next two sections, I consider why the preference for instrumental divergence, in particular, might be important for prosocial behavior.

3.4.2 The Cost of Instrumental Control

Theories of instrumental behavior distinguish between goal-directed and habitual action selection – whereas the former is flexible but computationally costly, the latter is "fast & frugal" (Liljeholm & O'Doherty, 2011). Supporting evidence can be found, for instance, in the fact that stress promotes habitual behavior at the expense of goal-directed behavior (Schwabe & Wolf, 2009), and in that working memory capacity protects from this stress response (Otto et al., 2013). Importantly, Forbes et al. (2023) assessed how stress influences participants' willingness to exert low-cost effort when earning a reward for themselves vs. others. They found that, compared to a control group, participants in the stress group favored putting in effort when gaining rewards for themselves vs. others. In the present study, an unwillingness to exert undue effort to benefit others might explain the difference in preference for high instrumental divergence when choosing for oneself vs. a charity.

3.4.3 Empathy vs. Reward

Spaans, Peters, & Crone (2019) found elevated activity in the ventral striatum – which functions include processing reward information (Schultz, 2016) – when gaining money for self and for both self and charity, but not when the charity was the only one to receive monetary gains. However, they found that activity in the ventral striatum – specifically in the nucleus accumbens – for charity gains was correlated with individual differences in empathic concern. Moreover, Spaans, Peters, & Crone (2019) suggest activity in the ventral striatum when observing gains for others is influenced by the strength of the personal connection to those others. In relation to the current study, these findings could be interpreted in at least two ways. Lack of activity in the ventral striatum during charity gains might reflect this type of gain is less rewarding. In general, people are less willing to engage in more costly activities if they produce lower rewards, so participants might not have been willing to put as much effort in gaining money for charities as in gaining money for themselves, and thus based their behavior on habitual action selection. Another possibility is, partially influenced by the lack of presentation and description of any charities (participants were simply told they would select a charity at the end of the study and that potential earnings would go to that charity, but did not read about any specific charity), thus reducing the probability of eliciting empathic concern in the participants, that a lack of activity in the ventral striatum lead to participants not engaging in more computationally intensive action selection. Support of this can be found in the review by Floresco (2015), which suggests that the nucleus accumbens plays a key role in action selection that facilitates goal-directed behavior, and its involvement might be more prominent when there is ambiguity or uncertainty about the best course of action.

3.4.5 Conclusion

In conclusion, the clear preference for environments with high instrumental divergence that has been shown when choices directly influence the decision maker, is not shown for choices that influence charities. The reason might be a higher reliance on less effortful, heuristic-based, habitual action selection when the outcomes have less benefits for the decision maker.

SUMMARY AND CONCLUSIONS

In this dissertation, I studied core cognitive and affective factors influencing charity donations. I focused on factors that have not been previously studied in direct relation to charitability, or that have been studied using a different approach; my general approach consisted of real-time interactive charitable donations of time and effort. This work adds relevant empirical knowledge to the prosocial behavior literature.

In Chapter 1, I assessed the influence of Narcissism & Social Approval Seeking on public vs. anonymous donations, and how social information about the donation decisions of others modulate those relationships. While Simulated Compassion increased the probability of donating when no social information was present, this effect was counteracted when people learned that the majority chose to make donations anonymously. What is more, higher Simulated Compassion predicted a decrease in the amount of help. In regards to Social Anxiety, a tendency to increase the probability of making a donation was found when social information about other people's decisions was present. For Narcissism, no evidence was found of this trait promoting prosocial behavior. In turn, Narcissism predicted a decrease in Anonymous donations, which were more beneficial to the charity than public donations.

In Chapter 2, I focused on self-conscious emotions. Specifically, I used an interpersonal guilt induction task to assess how self-conscious emotions in one domain influence charitable giving in another domain, again looking at the role of social information and assessing the probability and magnitude of donations. I additionally included a measure of the self-conscious traits of shame proneness, guilt proneness, externalization, and detachment. I found that participants that performed worse than the norm showed a tendency to have a

higher probability of donating than those who performed better than the norm, but only if they were not informed about what the social norm performance was. Participants that performed worse than the norm had a lower probability of donating when they learned about the social norm, than when they did not know how others performed. I interpreted this finding as showing a change in the emotion driving the behavior. When learning about the social norm, feelings of guilt were exceeded by feelings of shame, which has been related to finding oneself to be incompetent relative to established norms (Abell & Gecas, 1997). Shame's focus is the self (Tangney, 1990), and there is some evidence of shame reducing prosocial behavior (Ibanez & Roussel, 2021). Regarding self-conscious traits, when there was no direct manipulation of emotions, guilt proneness was correlated with an increase in the probability of donating, while shame proneness and externalization were associated with a decrease. In the case of detachment, when there was no direct manipulation of emotions, results showed a positive correlation with the magnitude of donations.

Finally, in Chapter 3, I looked at how a recently demonstrated preference for environments with high levels of instrumental control might differ across earning for oneself vs. a charity. I found that the preference for environments with high instrumental divergence was weakened – or absent – when decisions did not have an influence on oneself. Results suggested a preference for free choice when choosing for oneself more than when choosing for others. Although free choice is generally preferred to forced choice, it requires additional cognitive processes (e.g., Naefgen, Dambacher, & Janczyk, 2018). This additional cost has a higher probability of being rewarded in situations in which the outcomes can generate a direct benefit for the person making the effort, than when benefits go to charity. When earnings were for charity, participants showed a preference for higher outcome divergence,

even if there was no flexible control. These decisions may have been driven by a diversity heuristic, in which the higher the perceived diversity the lower the risk (Ayal & Zakay, 2009). Participants might have based their decisions on a heuristic that biased their decisions towards higher perceived diversity, rather than on more effortful computations. An unwillingness to exert undue effort to benefit others might explain the difference in preference for high instrumental divergence when choosing for oneself vs. a charity. People might have a higher reliance on less effortful, heuristic-based, habitual action selection when the outcomes have less benefits for themselves.

REFERENCES

Abell, E., & Gecas, V. (1997). Guilt, shame, and family socialization: A retrospective study. *Journal of Family Issues*, *18*(2), 99-123.

Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What does the narcissistic personality inventory really measure?. *Assessment*, *18*(1), 67-87.

Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics*, *92*(5-6), 1047-1060.

Ayal, S., & Zakay, D. (2009). The perceived diversity heuristic: the case of pseudodiversity. *Journal of personality and social psychology*, *96*(3), 559.

Bown, N. J., Read, D., & Summers, B. (2003). The lure of choice. *Journal of Behavioral Decision Making*, *16*(4), 297-308.

Bujang, M. A., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World*, *3*(1), 37-46.

Catarino, F., Gilbert, P., MCewaN, K., & Baião, R. (2014). Compassion motivations: Distinguishing submissive compassion from genuine compassion and its association with shame, submissive behavior, depression, anxiety and stress. *Journal of Social and Clinical Psychology*, *33*(5), 399-412.

Chapman, C. M., Louis, W. R., Masser, B. M., & Thomas, E. F. (2022). Charitable Triad Theory: How donors, beneficiaries, and fundraisers influence charitable giving. *Psychology & Marketing*, *39*(9), 1826-1848.

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*(39), 12315-12320.

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science*, *21*(5), 355-359.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of personality and social psychology*, *100*(5), 947.

Colwill, R. M., & Rescorla, R. A. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*(4), 520.

De Hooge, I. E. (2012). The exemplary social emotion guilt: Not so relationship-oriented when another person repairs for you. *Cognition & emotion*, *26*(7), 1189-1207.

De Hooge, I. E., Breugelmans, S. M., & Zeelenberg, M. (2008). Not so ugly after all: when shame acts as a commitment device. *Journal of personality and social psychology*, *95*(4), 933.

De Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition and emotion*, *21*(5), 1025-1042.

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. A. (2017). *The social psychology of prosocial behavior*. Psychology Press.

Floresco, S. B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annual review of psychology*, *66*, 25-52.

Forbes, P., Aydogan, G., Braunstein, J. T., Todorova, B., Wagner, I., Lockwood, P., Apps, M., Ruff, C., & Lamm, C. (2023). Acute stress reduces effortful prosocial behaviour.*eLife* 12:RP87271.

Gentile, B., Miller, J. D., Hoffman, B. J., Reidy, D. E., Zeichner, A., & Campbell, W. K. (2013). A test of two brief measures of grandiose narcissism: The Narcissistic Personality Inventory–13 and the Narcissistic Personality Inventory-16. *Psychological assessment*, *25*(4), 1120.

Gilbert, P., Catarino, F., Duarte, C., Matos, M., Kolts, R., Stubbs, J., ... & Basran, J. (2017a). The development of compassionate engagement and action scales for self and others. *Journal of Compassionate Health Care*, *4*, 1-24.

Gilbert, P., Catarino, F., Sousa, J., Ceresatto, L., Moore, R., & Basran, J. (2017b). Measuring competitive self-focus perspective taking, submissive compassion and compassion goals. *Journal of Compassionate Health Care*, *4*(1), 1-9.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: an evolutionary analysis and empirical review. *Psychological bulletin*, *136*(3), 351.

Greening, S., Norton, L., Virani, K., Ty, A., Mitchell, D., & Finger, E. (2014). Individual differences in the anterior insula are associated with the likelihood of financially helping versus harming others. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 266-277.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, *30*(2), 583-590.

Harris, N. (2019). *Shame, ethical identity and conformity: Lessons from research on the psychology of social influence*. The Australian National University, Regulatory Institutions Network (RegNet).

Hibbert, S., Smith, A., Davies, A., & Ireland, F. (2007). Guilt appeals: Persuasion knowledge and charitable giving. *Psychology & Marketing*, *24*(8), 723-742.

Ibanez, L., & Roussel, S. (2021). The effects of induced emotions on environmental preferences and behavior: An experimental study. *Plos one*, *16*(9), e0258045.

JASP Team. (2023). JASP (Version 0.17) [*Computer software*]. Retrieved from https://jasp-stats.org/

Jiang, Y., Bong, M., & Kim, S. I. (2015). Conformity of Korean adolescents in their perceptions of social relationships and academic motivation. *Learning and Individual Differences*, *40*, 41-54.

Ketelaar, T., & Tung Au, W. (2003). The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and emotion*, *17*(3), 429-453.

Łakuta, P. (2018). Social anxiety questionnaire (SAQ): Development and preliminary validation. *Journal of affective disorders*, *238*, 233-243.

Leary, M. R., & Kowalski, R. M. (1997). *Social anxiety*. Guilford Press.

Leith, K. P., & Baumeister, R. F. (1998). Empathy, shame, guilt, and narratives of interpersonal conflicts: Guilt-prone people are better at perspective taking. *Journal of personality*, *66*(1), 1-37.

Liljeholm, M. (2018). Instrumental divergence and goal-directed choice. In *Goal-Directed Decision Making* (pp. 27-48). Academic Press.

Liljeholm, M. (2021). Agency and goal-directed choice. *Current Opinion in Behavioral Sciences*, *41*, 78-84.

Liljeholm, M. (2022). Flexible control as surrogate reward or dynamic reward maximization. *Cognition*, *229*, 105262.

Liljeholm, M., & O'Doherty, J. P. (2011). Subcortical contributions to the motivational and cognitive control of instrumental performance by Pavlovian and discriminative stimuli. *Neural basis of motivational and cognitive control. Cambridge (Massachusetts): The MIT Press*, 149-162.

Liljeholm, M., Dunne, S., & O'Doherty, J. P. (2015). Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control. *European Journal of Neuroscience*, *41*(10), 1358-1371.

Liljeholm, M., Mistry, P., & Koh, S. (2018). The influence of schizotypal traits on the preference for high instrumental divergence. *Cog Sci 2018 Proceedings*.

Liljeholm, M., Wang, S., Zhang, J., & O'Doherty, J. P. (2013). Neural correlates of the divergence of instrumental probability distributions. *Journal of Neuroscience*, *33*(30), 12519-12527.

Lim, D., & DeSteno, D. (2016). Suffering and compassion: The links among adverse life experiences, empathy, compassion, and prosocial behavior. *Emotion*, *16*(2), 175.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory*, *37*(1), 145-151.

Miller, J. D., Gentile, B., Wilson, L., & Campbell, W. K. (2013). Grandiose and vulnerable narcissism and the DSM–5 pathological personality trait model. *Journal of personality assessment*, *95*(3), 284-290.

Miller, J. D., Lynam, D. R., Hyatt, C. S., & Campbell, W. K. (2017). Controversies in narcissism. *Annual review of clinical psychology*, *13*, 291-315.

Mistry, P., & Liljeholm, M. (2015). The Role of Outcome Divergence in Goal-Directed Choice. In *CogSci*.

Mistry, P., & Liljeholm, M. (2016). Instrumental divergence and the value of control. *Scientific reports*, *6*(1), 36295.

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the dark triad (narcissism, Machiavellianism, and psychopathy). *Perspectives on psychological science*, *12*(2), 183-204.

Naefgen, C., Dambacher, M., & Janczyk, M. (2018). Why free choices take longer than forced choices: Evidence from response threshold manipulations. *Psychological Research*, *82*, 1039-1052.

Nelissen, R. (2014). Relational utility as a moderator of guilt in social interactions. *Journal of Personality and Social Psychology*, *106*(2), 257.

Nelissen, R. M., Dijker, A. J., & deVries, N. K. (2007). How to turn a hawk into a dove and vice versa: Interactions between emotions and goals in a give-some dilemma game. *Journal of Experimental Social Psychology*, *43*(2), 280-286.

Nelissen, R., & Zeelenberg, M. (2009). When guilt evokes self-punishment: evidence for the existence of a Dobby Effect. *Emotion*, *9*(1), 118.

Norton, K. G., & Liljeholm, M. (2020). The rostrolateral prefrontal cortex mediates a preference for high-agency environments. *Journal of Neuroscience*, *40*(22), 4401-4409.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941-20946.

Pearce, P. L., & Amato, P. R. (1980). A taxonomy of helping: A multidimensional scaling analysis. *Social Psychology Quarterly*, 363-371.

Polman, E., & Ruttan, R. L. (2012). Effects of anger, guilt, and envy on moral hypocrisy. *Personality and Social Psychology Bulletin*, *38*(1), 129-139.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of personality and social psychology*, *54*(5), 890.

Ruggero G. Bettinardi (2023). computeCohen_d(x1, x2, varargin) (https://www.mathworks.com/matlabcentral/fileexchange/62957-computecohen_d-x1-x2-varargin), MATLAB Central File Exchange. Retrieved May 8, 2023.

Sasaki, S. (2019). Majority size and conformity behavior in charitable giving: Field evidence from a donation-based crowdfunding platform in Japan. *Journal of Economic Psychology*, *70*, 36-51.

Schmader, T., & Lickel, B. (2006). The approach and avoidance function of guilt and shame emotions: Comparing reactions to self-caused and other-caused wrongdoing. Motivation and Emotion, 30, 42-55.

Schultz, W. (2016). Reward functions of the basal ganglia. *Journal of neural transmission*, *123*, 679-693.

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, *29*(22), 7191-7198.

Silfver, M., Helkama, K., Lönnqvist, J. E., & Verkasalo, M. (2008). The relation between value priorities and proneness to guilt, shame, and empathy. *Motivation and emotion*, *32*, 69-80.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, *24*(18), R875-R878.

Spaans, J. P., Peters, S., & Crone, E. A. (2019). Neural reward-related reactions to monetary gains for self and charity. *Cognitive, Affective, & Behavioral Neuroscience*, *19*, 845-858.

Tangney, J. P. (1990). Assessing individual differences in proneness to shame and guilt: development of the Self-Conscious Affect and Attribution Inventory. Journal of personality and social psychology, 59(1), 102.

Tangney, J. P., Dearing, R. L., Wagner, P. E., & Gramzow, R. (2000). *Test of Self-Conscious Affect–3 (TOSCA-3)*.

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions?. Journal of personality and social psychology, 70(6), 1256.

Teachman, B. A., & Allen, J. P. (2007). Development of social anxiety: Social interaction predictors of implicit and explicit fear of negative evaluation. *Journal of Abnormal Child Psychology*, *35*, 63-78.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). *Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)*

Treeby, M., & Bruno, R. (2012). Shame and guilt-proneness: Divergent implications for problematic alcohol use and drinking to cope with anxiety and depression symptomatology. *Personality and Individual Differences*, *53*(5), 613-617.

Ty, A., Mitchell, D. G., & Finger, E. (2017). Making amends: Neural systems supporting donation decisions prompting guilt and restitution. *Personality and Individual Differences*, *107*, 28-36.

Uji, M., Kitamura, T., & Nagata, T. (2011). Self-conscious affects: Their adaptive functions and relationship to depressive mood. *American Journal of Psychotherapy*, *65*(1), 27-46.

Weaver, R.; Frederick, S. (2012). "A Reference Price Theory of the Endowment Effect". *Journal of Marketing Research*. **49** (5): 696–707.

Weisman, O., Aderka, I. M., Marom, S., Hermesh, H., & Gilboa-Schechtman, E. (2011). Social rank and affiliation in social anxiety disorder. *Behaviour research and therapy*, *49*(6-7), 399-405.

Yu, H., Hu, J., Hu, L., & Zhou, X. (2014). The voice of conscience: neural bases of interpersonal guilt and compensation. Social Cognitive and Affective Neuroscience, 9(8), 1150-1158.

Zagefka, H., & James, T. (2015). The psychology of charitable donations to disaster victims and beyond. *Social Issues and Policy Review*, *9*(1), 155-192.

Zhu, R., Feng, C., Zhang, S., Mai, X., & Liu, C. (2019). Differentiating guilt and shame in an interpersonal context with univariate activation and multivariate pattern analyses. *NeuroImage*, *186*, 476-486.